

# The stellar correlation function from SDSS

## A statistical search for wide binary stars

M. Longhitano and B. Binggeli

Department of Physics, Basel University, Klingelbergstrasse 82, CH-4056 Basel  
e-mail: marco.longhitano@unibas.ch

Received 12 August 2009 / Accepted 1 October 2009

### ABSTRACT

We study the statistical properties of the wide binary population in the Galaxy field with projected separations larger than 200 AU by constructing the stellar *angular* two-point correlation function (2PCF) from a homogeneous sample of nearly 670 000 main sequence stars. The selected stars lie within a rectangular region around the Northern Galactic Pole and have apparent *r*-band magnitudes between 15 and 20.5 mag and spectral classes later than G5 ( $g - r > 0.5$  mag). The data were taken from the Sixth Data Release of the Sloan Digital Sky Survey. We model the 2PCF by means of the Wasserman-Weinberg technique including several assumptions on the distribution of the binaries' orbital parameters, luminosity function, and density distribution in the Galaxy. In particular, we assume that the semi-major axis distribution is described by a single powerlaw. The free model parameters – the local wide binary number density  $n_{\text{WB}}$  and the power-law index  $\lambda$  of the semi-major axis distribution – are inferred simultaneously by least-square fitting. We find the separation distribution to follow Öpik's law ( $\lambda = 1$ ) up to the Galactic tidal limit, without any break and a local density of 5 wide binaries per 1 000 pc<sup>3</sup> with both components having spectral type later than G5. This implies that about 10% of all stars in the solar neighbourhood are members of such a late-type wide binary system. With a relative statistical ( $2\sigma$ ) error of about 10%, our findings are in general agreement with previous studies of wide binaries. The data suggest that about 800 very wide pairs with projected separations larger than 0.1 pc exist in our sample, whereas none are found beyond 0.8 pc. Modern large-scale surveys make the 2PCF method a viable tool for studying wide binary stars and a true complement to common proper motion studies. The method is, however, seriously limited by the noise from optical pairs and the (over)simplifying assumptions made to model the selection effects and to interpret the measured clustering signal.

**Key words.** Binaries: general – Binaries: visual – Galaxy: stellar content – Stars: statistics

### 1. Introduction

Binary stars have traditionally been central for astronomy, especially *close* binary systems (with typical orbital periods of days to years), because they are genuine laboratories of stellar evolution and its exotic remnants, making them a cornerstone for determining masses, distances, and many other fundamental astrophysical parameters. On the other hand, *wide* binary systems (orbital periods of many thousands to millions of years) are of particular interest, too.

Because they are only weakly bound by gravitation, wide binaries are prone to tidal disruption by passing massive objects, such as massive stars, molecular clouds, MACHOs (massive compact halo objects), or dark matter (DM) substructure. The shape of their separation distribution, in particular towards the most extreme, widest separations, should therefore allow constraints on the mass and frequency of the disruptive perturbers (Ritterer & King 1982), as well as to estimate the age of a population (Poveda & Allen 2004). Wide binary-based MACHO constraints have been placed (but subsequently criticised) by several authors (Bahcall et al. 1985a; Weinberg et al. 1987; Weinberg 1990; Wasserman & Weinberg 1991; Yoo et al. 2004; Quinn et al. 2009).

In a different context, Hernandez & Lee (2008) propose using the tightening of wide binaries in dwarf spheroidal galaxies through dynamical friction as a test for DM. Moreover, binaries with separations over 0.1 pc, which are known to exist in the Galaxy (e.g. Lépine & Bongiorno 2007, also this paper),

are in the ‘weak-acceleration’ regime where, in principle, one could test for possible deviations from Newtonian gravity, such as MOND (Milgrom & Bekenstein 1987; Close et al. 1990).

Wide binaries are not only probes of dynamical evolution and Galactic structure, but they also provide important clues to star formation (Larson 2001). The exact outcome of the binary population in a star-formation event is an exceedingly complex and still unsolved problem (Goodwin et al. 2007, and references therein). The formation of extremely wide binaries is particularly difficult to understand (Allen et al. 2007; Parker et al. 2009; Kouwenhoven et al. 2010). A good knowledge of the wide binary frequency and separation distribution is primary for a whole host of problems in astrophysics (e.g. Chanamé 2007).

Unfortunately, the long orbital periods of wide binary systems make their identification very difficult in the first place. There are basically two different methods for detecting a wide binary system: (1) by the number excess of neighbours around a given star with respect to a random distribution, or (2) by the common proper motion (CPM) of two well-separated stars. Although there is no way to observe the orbital motion of a wide binary pair, the CPM is nevertheless a reliable indicator of a physical relationship between two individual stars (e.g. Lépine & Bongiorno 2007). In the present study we use the *angular two-point correlation function* (2PCF) to measure the excess of neighbours compared to a random distribution. This method has the advantage that larger samples of more distant

stars can be used, as only the stars' positions are required. It only allows, however, statistical statements on the pairing and is limited to relatively small angular separations, because the noise due to randomly associated pairs increases rapidly (linearly) with angular separation.

To date, our knowledge on wide binaries essentially had rested on studies using the CPM method. There are two excellent recent CPM studies of wide binaries by Chanamé & Gould (2004) and Lépine & Bongiorno (2007). Based on updated proper motion (Luyten and Hipparcos) catalogues of stars, 917 and 521 wide binary systems, respectively, have been identified over the entire Northern sky and with a typical median distance of about 100 pc. Lépine & Bongiorno found the separation distribution of the pairs to follow Öpik's (1924) law, i.e. frequency being proportional to the inverse of separation, out to separations of around 3 500 AU. Beyond this characteristic scale, however, the separation distribution seems to be falling by a steeper power law, without an obvious cut-off. Lépine & Bongiorno find wide binaries out to separations of almost 100 000 AU, or 0.4 pc! They quote a number of 9.5% for nearby ( $D < 100$  pc) Hipparcos stars belonging to a wide binary system with a separation greater than 1 000 AU, again demonstrating the ubiquity of the (wide) binary phenomenon. Chanamé & Gould, in addition, achieve a distinction between wide binaries belonging to the Galactic disc and those belonging to the Galactic halo. No significant difference in the separation distribution has been found, suggesting that the disc and halo binaries were formed under similar conditions, despite the very different metallicities and ages.

The angular 2PCF (also called "covariance function") is one of the most useful tools for studying the clustering properties of galaxies, and it has been widely used to probe the large-scale structure of the Universe (e.g. Peebles 1980, for a more recent study of the angular clustering of galaxies see Scranton et al. 2002 and references therein.). The 2PCF can also be, and has been, used to measure the clustering properties of stars. While on large scales the stars are clearly randomly distributed in the Galaxy, as expected in a collisionless system, correlations up to the 10 pc scale (or even beyond) can be found in moving groups and halo streamers (Doinidis & Beers 1989), star-forming regions (e.g. Gomez & Lada 1998), and open star clusters (López-Corredoira et al. 1998). On very small scales (sub-pc, observationally: sub-arcmin), there is a strong signal due to (visual) binary stars.

The 2PCF method of probing wide binary stars in the Galactic field was pioneered by Bahcall & Soneira (1981), who studied the distribution of stars down to a limiting magnitude  $V = 16$  in a 10 square degree field at the NGP, and found very significant clustering at a separation of 0.1 pc. Of the 19 binary candidates, 6 turned out to be real (Latham et al. 1984). The theoretical implications of these observations for the frequency and separation distribution of the binaries, and a general method for modelling them, was worked out by Wasserman & Weinberg (1987). To date, there are only few follow-up studies of the stellar 2PCF (Saarinen & Gilmore 1989; Garnavich 1988, 1991; Gould et al. 1995). Given this surprising lack of further work on the stellar 2PCF to study wide binaries, and in view of the enormous progress in deep photometric sky surveys in the past 20 years that should render the 2PCF method – as a true complement to the CPM method – much more rewarding now, we have started a project to use the huge stellar database of the Sloan Digital Sky Survey (SDSS York et al. 2000, [www.sdss.org](http://www.sdss.org)) for a detailed stellar correlation analysis to very faint magnitudes.

An independent study of wide binaries in the SDSS database by Sesar et al. (2008) takes a different approach, based on the

"Milky Way Tomography" of Jurić et al. (2008), where approximate distances are ascribed to all stars by adopting a photometric parallax relation. Candidate binaries are selected by the requirement that the difference between two potential components in apparent magnitude is within a certain error equal to the difference in absolute magnitude.

Using distance information has the fundamental advantage of filtering out the disturbing noise from chance projections and of avoiding the need for a complex, Wasserman & Weinberg-type modelling to calculate integrated, sky-projected quantities. Although we plan to include distance information in future work, we show here that the *angular* correlation analysis is in principle still a viable approach. Our main results agree with those of the CPM studies. We also show, however, what the limitations of the method are.

The paper is organised as follows. In Sect. 2 we describe the SDSS data input and define the cleansed sample used for the analysis. Section 3 is devoted to the angular correlation function apparatus, followed by an extensive description of the modelling of the correlation function drawing on a modified Wasserman & Weinberg technique in Sect. 4. Our results for the total sample and for a number of subsamples are presented in Sect. 5 and are critically discussed and compared to previous work in Sect. 6. Concluding remarks are given in Sect. 7.

## 2. Data

For our analysis we selected a homogeneous sample of stars from the Sixth SDSS Data Release (DR6; Adelman-McCarthy et al. 2008). We took a rectangular area around the Northern Galactic Pole (NGP) covering approximately  $\Omega_{\text{tot}} \simeq 675$  square degrees (dec:  $22^\circ - 42^\circ$ , RA:  $165^\circ - 205^\circ$ ; see Fig. 1). It contains 966 656 *primary*<sup>1</sup> point-like objects – including quasars, asteroids, and possibly some misidentified galaxies – selected from the *Star* view<sup>2</sup>, having an apparent PSF magnitude<sup>3</sup> in the *r*-band between 15 and 20.5 mag. Following the SDSS recommendations<sup>4</sup>, we required the stars to have *clean photometry*<sup>5</sup> in the *g*, *r*, and *i* band. Although the effect of interstellar dust on the measurements is weak in the direction of the NGP, we corrected the data using the Schlegel et al. (1998) maps, which can be easily done, as the extinction at the position of each object is stored in the SDSS database.

The magnitude limits were chosen to be within the saturation limit of the SDSS CCD cameras ( $r \sim 14$  mag; Gunn et al. 1998)<sup>6</sup> and the limit where the star-galaxy separation becomes unreliable ( $r \sim 21.5$  mag; Lupton et al. 2001)<sup>7</sup>. Additionally, we avoid close stars being overcorrected by the adopted extinction correction, since most stars with  $r > 15$  mag are behind the entire dust column (Jurić et al. 2008). On the other hand, with the somewhat conservative faint limit of  $r = 20.5$  we make sure that only very few stars have a magnitude in the *g* or *i* band beyond

<sup>1</sup> Due to overlaps in the imaging, many objects are observed more than once. The best of those observations is called the *primary object*, all the others are called *secondary objects*.

<sup>2</sup> See [cas.sdss.org/dr6/en/help/browser/browser.asp](http://cas.sdss.org/dr6/en/help/browser/browser.asp)

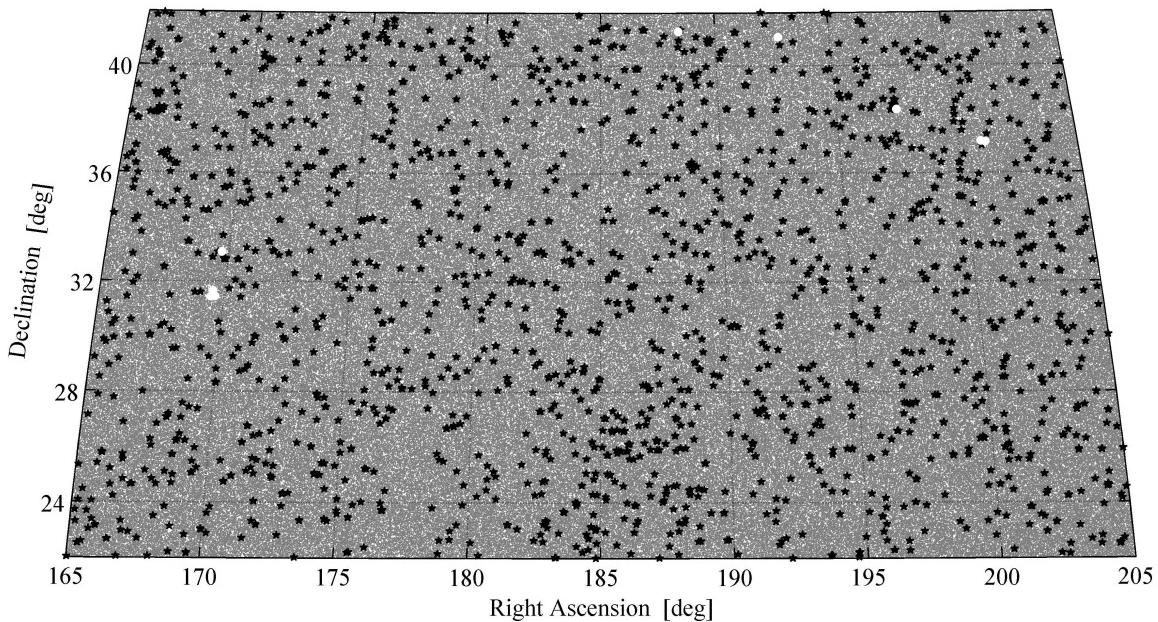
<sup>3</sup> See [sdss.org/dr6/algorithms/photometry.html#mag\\_psf](http://sdss.org/dr6/algorithms/photometry.html#mag_psf)

<sup>4</sup> See [sdss.org/dr6/products/catalogs/flags.html](http://sdss.org/dr6/products/catalogs/flags.html)

<sup>5</sup> See [cas.sdss.org/dr6/en/help/docs/realquery.asp#flags](http://cas.sdss.org/dr6/en/help/docs/realquery.asp#flags)

<sup>6</sup> See also [astro.princeton.edu/PBOOK/camera/camera.htm](http://astro.princeton.edu/PBOOK/camera/camera.htm)

<sup>7</sup> See also [sdss.org/dr6/products/general/stargalsep.html](http://sdss.org/dr6/products/general/stargalsep.html)



**Fig. 1.** Distribution of stars (grey points filling the background) in our total sample. Black asterisks show the positions of bright star masks, white circles those of hole masks. The size of the symbols are not to scale.

the 95% completeness limit of 22.2 mag or 21.3 mag, respectively (Adelman-McCarthy et al. 2007, Table 1)<sup>8</sup>.

The average seeing of the SDSS imaging data (median PSF width) is  $1.4''$  in the  $r$ -band (Adelman-McCarthy et al. 2008)<sup>9</sup>. To be on the safe side, we took the minimum angular separation to be  $\theta_{\min} = 2''$ .

### 2.1. Contaminations

Matching our sample with the *QsoBest* table<sup>10</sup> resulted in the exclusion of 10 041 quasar candidates. Most of them (8 157) have  $g - r \lesssim 0.5$  mag and are scattered in a colour-colour diagram around the otherwise narrow stellar locus as shown in Fig. 2. Even after removing the quasar candidates, the remaining objects in the blue part of our sample show a suspicious scatter, which is probably caused by further quasars and misidentified galaxies. We therefore decided to exclude all objects with  $g - r < 0.5$  mag, removing a further 286 227 objects from our sample. Furthermore, we reject all moving objects (asteroids) by cutting on the DEBLENDED\_AS\_MOVING flag. This leaves us with 670 388 objects classified as “stars” in the sample.

The large majority of the stars observed by the SDSS are main sequence (MS) stars. Finlator et al. (2000) estimate the fraction of stars that are not on the MS to be  $\sim 1.0\%$ , most of them giants and subgiants ( $\sim 90\%$ ), but also horizontal branch stars ( $\sim 10\%$ ). The number of white dwarfs observed by the SDSS is negligible compared to the number of MS stars (e.g. Harris et al. 2006). Thus, it seems well-justified to assume that *all* the stars in our sample are on the MS.

The cut discussed above at  $g - r = 0.5$  mag implies that our sample contains only stars from spectral type later than about G5 (Finlator et al. 2000). In addition, this cut is appropriate for our purposes for the following three reasons:

<sup>8</sup> See also [sdss.org/dr6/](http://sdss.org/dr6/)

<sup>9</sup> See also <http://www.sdss.org/dr6/> and the DR5 paper (Adelman-McCarthy et al. 2007)

<sup>10</sup> See [cas.sdss.org/dr6/en/help/docs/algorithms.asp?key=qsocat](http://cas.sdss.org/dr6/en/help/docs/algorithms.asp?key=qsocat)

*i)* Because they are very young, the bluest MS stars are mostly members of loose associations, so their clustering properties still represent the peculiarities of their birth places. Being an interesting subject to study (e.g. Kobulnicky & Fryer 2007), excluding them assures that our clustering signal is predominantly due to wide binaries in the field that have lost their memory of their birth places.

*ii)* As the MS becomes more sparsely populated towards the blue end, and a significant fraction could be made up by metal-poor halo giants, the assumption that all stars in our sample are on the MS might not be valid for the bluest stars.

*iii)* For magnitudes  $M_V \gtrsim 4.5$  mag, the shape of the halo luminosity function agrees well with that of the disc luminosity function (e.g. Bahcall et al. 1985b, their Fig. 2). Thus, the cut on  $g - r = 0.5$  mag, which corresponds to a cut at  $M_V \simeq 5.6$  mag, allows us to use the disc luminosity function for the halo component.

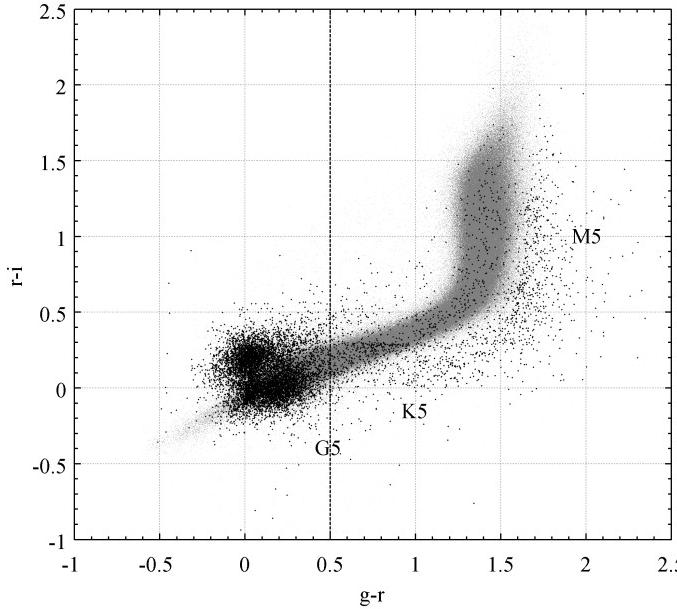
### 2.2. Survey holes and bright stars

The sample chosen contains some regions where, for various reasons, no object could be observed. These regions are masked<sup>11</sup> as *hole* in the SDSS database and can therefore be easily identified. Holes in the sample affect our analysis in two ways: first, they diminish the total area of the sample. Second, they introduce an edge effect, as stars near a hole show a lack of neighbouring stars. The former effect can be easily corrected for in an approximate way as the SDSS provides the radius  $\theta_M^{(i)}$  of its bounding circle for every mask  $i$ . The residual area of the sample is then

$$\Omega \simeq \Omega_{\text{tot}} - \sum_{i=1}^{N_M} \pi (\theta_M^{(i)})^2 \quad (1)$$

where  $N_M$  denotes the number of hole masks. We discuss the correction for edge effects in Sect. 3.2 in the context of the correlation function.

<sup>11</sup> See [sdss.org/dr6/algorithms/masks.html](http://sdss.org/dr6/algorithms/masks.html)



**Fig. 2.** Colour-colour diagram of all point-like objects within our coordinate's and apparent magnitude's limits (grey points). Black points are SDSS quasar candidates. The vertical dashed line indicates the cut at  $g - r = 0.5$  mag (see text). Approximate spectral classes are indicated.

Restricting ourselves to the masks defined in the  $r$ -band, we find  $N_{\text{SH}} = 9$  regions masked as “hole” in our sample (survey holes: SH) with an average radius of  $\bar{\theta}_{\text{SH}} \approx 7.76$  arcmin. These survey holes diminish the total area of the sample by approximately 0.46 square degrees, constituting a marginal correction that could be safely ignored.

Very bright (saturated) stars cause similar problems: a very bright object may appear like a hole in our sample, when the underlying fainter stars blended with this object cannot be revealed. Within our sample there are  $N_{\text{BS}} = 1790$  *bright star* (BS) masks in the  $r$ -band. We exclude all objects inside such a mask, removing 545 stars from our sample. The bounding circles of bright star masks have an average radius of  $\bar{\theta}_{\text{BS}} \approx 1.95$  arcmin, resulting in a further diminishing of the total area of the sample of circa 5.96 square degrees. By correcting the sample's total area for both hole and bright star masks, we get a residual area  $\Omega \approx 675 - 0.46 - 5.96 \text{ deg}^2 = 668.58 \text{ deg}^2$ .

### 2.3. Final sample

Our final sample contains  $N_{\text{obs}} = 669\,843$  MS stars with *clean photometry*, a spectral type later than about G5 ( $g - r \geq 0.5$  mag), and an apparent magnitude in the range  $15 \leq r \leq 20.5$  mag. The stars are distributed over a solid angle  $\Omega \approx 668.58$  square degrees (after correcting for bright star masks and minor survey holes), corresponding to a mean surface density of  $n = N_{\text{obs}}/\Omega \approx 1\,000$  stars per square degree. Throughout this study, we use the terms “total sample” and “final sample” synonymously.

## 3. Stellar correlation function

### 3.1. Estimation of the correlation function

The angular two-point *auto*-correlation function (2PCF)  $w(\theta)$  is defined via the joint probability  $d\mathcal{P}$  of finding objects in both the solid angles  $d\Omega_1$  and  $d\Omega_2$

$$d\mathcal{P} = n^2 (1 + w(\theta)) d\Omega_1 d\Omega_2, \quad (2)$$

where  $n$  is the mean density of objects in the sky and  $\theta$  is the separation between the two areas (e.g. Peebles 1980, §45). The number of distinct stellar pairs,  $F(\theta)d\theta$ , with an angular separation between  $\theta$  and  $\theta + d\theta$ , *observed* in a region of angular size  $\Omega$ , can then be written as

$$F(\theta)d\theta = \frac{N_{\text{obs}}(N_{\text{obs}} - 1)}{2} (1 + w(\theta)) \frac{d\Omega(\theta)}{\Omega} \quad (3)$$

where  $N_{\text{obs}}$  is the number of stars observed in  $\Omega$  and  $d\Omega(\theta)$  is approximately equal to the solid angle of a ring with (middle) radius  $\theta$  and width  $d\theta$

$$d\Omega(\theta) \simeq 2\pi\theta d\theta. \quad (4)$$

Solving for  $w$  in Eq. 3 yields a simple estimator for the 2PCF:

$$\hat{w}(\theta) = \frac{F(\theta)}{P(\theta)} - 1, \quad (5)$$

where  $P(\theta)$  is the number of pairs expected from a random sample with  $N_{\text{obs}}$  points ( $w = 0$  in Eq. 3):

$$P(\theta)d\theta \simeq \frac{\pi N_{\text{obs}}(N_{\text{obs}} - 1)}{\Omega} \theta d\theta. \quad (6)$$

The 2PCF estimate  $\hat{w}(\theta)$  is a measure for the excess of observed pairs separated by an angle  $\theta$  with respect to a randomly distributed sample.

Another statistical measure equivalent to the 2PCF, which we use for visualising the data, is the *cumulative difference distribution* (CDD)  $\hat{\gamma}(\theta)$ , giving the cumulative number of pairs in excess of a random distribution:

$$\hat{\gamma}(\theta) \equiv \int_{\theta_{\min}}^{\theta} (F(\theta') - P(\theta')) d\theta' = \int_{\theta_{\min}}^{\theta} \hat{w}(\theta') P(\theta') d\theta', \quad (7)$$

with  $\theta_{\min} = 2''$  (see above). The CDD is completely equivalent to the 2PCF, but is – as for all cumulative distributions – more sensitive to statistical trends caused over a wider range of angular separations (i.e. over several bins), at the expense of a strong correlation of its values at different separations. The use of the simple estimator in Eq. 5 is appropriate as long as boundary effects due to the finite sample size are negligible (Bernstein 1994). We discuss boundary effect in the next section. The number of pairs observed  $F(\theta)$  is determined very efficiently *inside* the SDSS database using the precalculated *Neighbors* table which contains all the objects within  $\theta_{\max} = 30''$  of any given object.

### 3.2. Boundary effects

Stars close to the sample's boundary have a somewhat truncated ring  $d\Omega(\theta)$  and consequently show a lack of neighbours. However, as in the present study the probed angular scale is small compared to the sample's size, it turns out that these edge effects have a negligible impact on our results: we estimate the relative error introduced in  $d\Omega(\theta)$  by omitting the edge correction to be less than  $\sim 0.04\%$  for any given  $\theta \leq 30''$ .

The stars near the boundary of a hole or a bright star mask (in the following simply ‘‘hole’’) show a lack of neighbours, too. Therefore, only a fraction  $F_H^{\text{tot}}(\theta)$  of all pairs separated by an angular distance  $\theta$  has been observed, and we need to correct the observed number of pairs  $F(\theta)$ . The ‘‘true’’ number of pairs, corrected for edge effects, then reads as

$$F_{\text{corr}}(\theta) = \frac{F(\theta)}{F_H^{\text{tot}}(\theta)}. \quad (8)$$

In calculating the fraction  $F_H^{\text{tot}}$  we proceeded in the same way as López-Corredoira et al. (1998). The calculation is outlined in Appendix A. We find that the relative error in  $F$  for  $\theta = 30''$  amounts to

$$\Delta_F \equiv \frac{F_{\text{corr}} - F}{F_{\text{corr}}} = 1 - F_H^{\text{tot}} \simeq 0.7\%. \quad (9)$$

For smaller angular separations, the effect is even less.

Taking into consideration these edge effects, we rewrite the 2PCF estimate

$$\hat{w}_{\text{corr}}(\theta) = \frac{F_{\text{corr}}(\theta)}{P(\theta)} - 1 \quad (10)$$

and the CDD

$$\hat{\gamma}_{\text{corr}}(\theta) = \int_{\theta_{\min}}^{\theta} \hat{w}_{\text{corr}}(\theta') P(\theta') d\theta'. \quad (11)$$

### 3.3. Uncertainty of the correlation function estimate

As the values of  $\hat{w}_{\text{corr}}$  at different separations are *not* independent, Poisson errors may underestimate the uncertainty in  $\hat{w}_{\text{corr}}$ , especially when the angular scales under consideration are large (Hamilton 1993; Bernstein 1994). However, as we see in Sect. 5, the clustering of wide binary stars occurs on small angular scales ( $\theta \leq 15''$ ) compared to the size of our sample (20 deg  $\times$  40 deg). We therefore expect that using Poisson errors only underestimates the true errors by a small amount in our case. Thus, we adopt Poissonian errors on  $F(\theta)$ :

$$\delta F(\theta) = \sqrt{F(\theta)} \quad \text{and} \quad \delta F_{\text{corr}}(\theta) = \sqrt{\frac{F_{\text{corr}}(\theta)}{F_H^{\text{tot}}(\theta)}}. \quad (12)$$

Using Eq. 10 and Gauss’ error propagation formula, we may write the uncertainty of the CF estimate  $\hat{w}_{\text{corr}}$  as

$$\delta \hat{w}_{\text{corr}} = \frac{\delta F_{\text{corr}}}{P} = \sqrt{\frac{\hat{w}_{\text{corr}} + 1}{F_H^{\text{tot}} P}}, \quad (13)$$

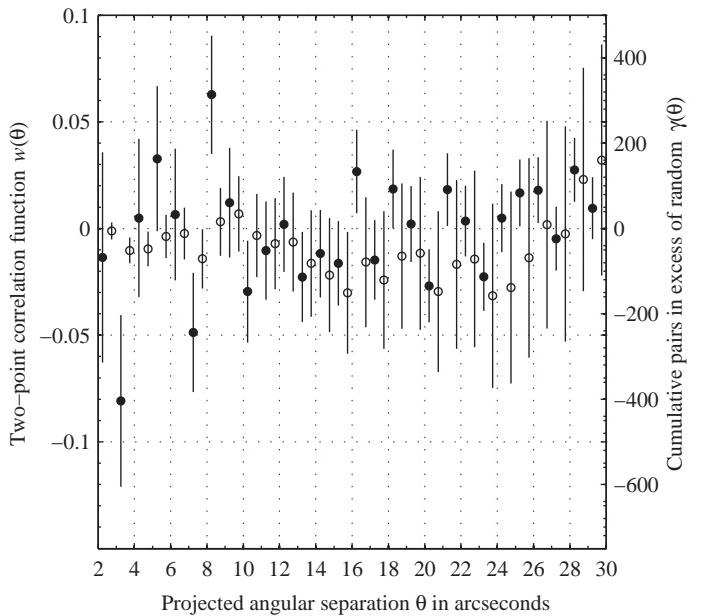
where we omitted the dependencies on  $\theta$  for the sake of brevity.

The uncertainty in the CDD  $\hat{\gamma}$  is easily obtained in the same way:

$$\delta \hat{\gamma}_{\text{corr}}(\theta) = \sqrt{\int_{\theta_{\min}}^{\theta} \delta F_{\text{corr}}^2(\theta') d\theta'}. \quad (14)$$

### 3.4. Testing the procedure for a random sample

To test the validity of the procedure to estimate the 2PCF described in the previous sections, we generated a random sample having the same number of ‘‘stars’’ and distributed over the same area as the stars in our final sample. In addition, we made certain



**Fig. 3.** 2PCF as inferred from a random sample (solid circles, left ordinate) and the corresponding CDD (open circles, right ordinate). Poisson errors are indicated as vertical lines.

that the random sample contains the same number of holes with appropriate radii.

For the analysis of the random sample we wrote a dedicated computer program that lays a grid with a mesh size of 1 arcmin over the sample before counting the pairs. This grid acts as a distance filter and avoids calculating the distances between all possible pairs: only pairs with both their components within a cell or with them in two adjacent cells, respectively, were taken into account.

If our procedure is correct we would expect a zero signal in both the 2PCF and the CDD when analysing a sample of randomly distributed stars. Figure 3 shows the result of the analysis of our random sample. There is indeed no clustering signal evident, neither in the 2PCF nor in the CDD. The results are consistent throughout with a zero signal out to the separation limit of 30'' (compare also with Fig. 5). This shows that our procedure for estimating the 2PCF is reliable.

## 4. The model

Our approach to modelling the angular 2PCF is based on a technique developed by Wasserman & Weinberg (1987), hereafter WW87. As we describe in more detail in the following sections, this technique makes some simple assumptions on the basic statistical properties of wide binaries, and projects these theoretical distributions on the observational plane using the selection criteria of a given (binary) star catalogue and the geometry of the Milky Way galaxy (see also Weinberg 1988).

Their long periods make it virtually impossible to distinguish wide binary stars from mere chance projections (optical pairs) by their orbital motion. Therefore, we do not attempt to identify individual wide binaries, but we look for a *statistical* signal stemming from physical wide pairs in the sample, solely exploiting precise stellar position measurements provided by the SDSS.

The original Wasserman & Weinberg (WW) technique calculates the projected separation distribution to compare it with a sample of wide binaries of known distance and angular sep-

eration. In the present study we are only dealing with angular separations. The calculation of wide binary counts as a function of angular separation alone requires an appropriate modification of the WW-technique (see also Garnavich 1991), which we discuss in Sect. 4.3.

#### 4.1. Wasserman-Weinberg technique

WW87 developed “a versatile technique for comparing wide binary observations with theoretical semimajor axis distributions”. According to WW87 we may write the number of observed binaries  $\psi(s)ds$  with projected physical separations between  $s$  and  $s + ds$  in a given catalogue of stars as

$$\psi(s)ds = n_{\text{WB}}Q(s)V(s)ds, \quad (15)$$

where  $n_{\text{WB}}$  is the total number density of wide binaries in the solar neighbourhood<sup>12</sup>,  $Q(s)$  is the reduced distribution of projected separations, and  $V(s)$  the “effective volume”.

The total number density  $n_{\text{WB}}$  is one of the two free parameters in the model that will be inferred by fitting the model to the observational data. The reduced separation distribution  $Q(s)$  contains the physical properties from the wide binaries (semimajor axis distribution, distribution of eccentricities, orientation of the orbital planes) projected onto the observational plane, whereas the effective volume  $V(s)$  takes into account the characteristics of the sample under consideration (covered area, range of angular separation examined, magnitude limits), as well as the stellar density distribution in the Galaxy and the luminosity function.

As we show in Sect. 4.3, this neat formal splitting in physical properties and selection effects will not be possible anymore after the required modification on the WW-technique mentioned above. At this point, we discuss the two parts –  $Q(s)$  and  $V(s)$  – more in detail.

##### 4.1.1. Reduced separation distribution

The reduced separation distribution  $Q(s)$  is given by the projection of the reduced (present-day) semi-major axis distribution  $q(a)$  against the sky. Gravitational perturbations due to stars, giant molecular clouds, and (hypothetical) DM particles cause the semi-major axes of (disc) wide binaries to evolve. Little is known about the initial semi-major axis distribution, and usually a single power-law is assumed, because of its simplicity (Weinberg 1988) but also because of theoretical considerations (Valtonen 1997; Poveda et al. 2007, and references therein). The evolution of the semi-major axis distribution of disc wide binaries has been modelled by Weinberg, Shapiro, & Wasserman (1987) in terms of the Fokker-Plack equation. Their numerical simulations suggest that the semi-major axis distribution evolves in a self-similar way for reasonable choices of the initial power-law index and the wide binary birth rate function. We decided, therefore, to model the present-epoch semi-major axis distribution of wide binary stars  $q(a)$  by a single powerlaw

$$q(a) \propto \left(\frac{a}{\text{pc}}\right)^{-\lambda} \text{pc}^{-1} \quad (16)$$

with the power-law index  $\lambda$ , the second free parameter of the model. The range in  $a$  where Eq. 16 is a valid description of the observed semi-major axis distribution at the same time provides the *definition* of what we consider as a “wide binary star”. We

specify this range by an upper and a lower limit,  $a_{\min}$  and  $a_{\max}$ , respectively. There is observational evidence that the distribution of the semi-major axis of binary stars has a break at 0.001 pc (Abt 1983). Following WW87 and Weinberg (1988), we thus take the lower limit, providing the division between the close and wide binary populations, to be  $a_{\min} = 0.001 \text{ pc} (\approx 200 \text{ AU})$ . On the other hand, the Galactic tides provide a natural maximum semi-major axis  $a_T$  beyond which no bound orbits can exist. Details on the calculation of  $a_T$  are given in Appendix B. We find it to be of the order of 1 pc.

We normalise  $q(a)$  – hence the name “reduced semi-major axis distribution” – such that

$$\int_{a_{\min}}^{a_T} q(a) da = 1 \quad (17)$$

giving

$$q(a) = c_\lambda \left(\frac{a}{\text{pc}}\right)^{-\lambda} \text{pc}^{-1} \quad (18)$$

where

$$c_\lambda = \begin{cases} (1 - \lambda) [a_T^{1-\lambda} - a_{\min}^{1-\lambda}]^{-1}, & \text{for } \lambda \neq 1 \\ [\ln a_T - \ln a_{\min}]^{-1}, & \text{for } \lambda = 1 \end{cases} \quad (19)$$

with  $a_{\min}$  and  $a_T$  in pc.

In a more general treatment,  $q(a)$  might also depend on the luminosity classes of the binary star members, as well as on their magnitudes. Like WW87, we restrict our analysis to models where  $q(a)$  only depends on the semi-major axis  $a$ .

It is somewhat disputed whether the semi-major axis distribution has a break at larger  $a$  attributed to the disruptive effects of the environment on the widest binary stars. In their extensive work Wasserman & Weinberg (1991) (see also Weinberg 1990) conclude that “although the data *suggest* a break in the physical distribution of wide binary separations, they do not *require* a break with overwhelming statistical significance”, whereas the more recent study by Lépine & Bongiorno (2007) shows evidence for a break at  $s \sim 3500 \text{ AU}$  (statistically, we have  $\langle s \rangle \simeq a$ ; see Eq. 22). In this context, the question arises whether the assumption that the data is described by a single powerlaw (Eq. 16) up to the tidal limit  $a_T$  can be rejected with confidence. We address this question in Sect. 5.

Assuming that “the binaries’ orbital planes are randomly oriented and that their eccentricities  $e$  are distributed uniformly in  $e^2$ ” (WW87) we can project  $q(a)$  against the sky by (Weinberg 1988)

$$Q(s) = \int_{s/2}^{\infty} \frac{da}{a} q(a) \mathcal{F}(s/a), \quad (20)$$

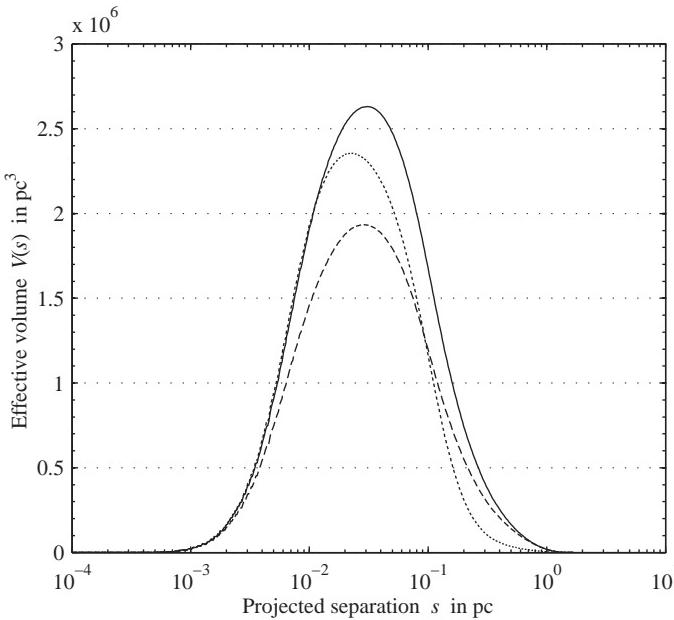
where  $\mathcal{F}(x)$  takes the eccentricity and angular averaging into account for the pairs, which may be written as (WW87, Weinberg & Wasserman 1988)

$$\mathcal{F}(x) = \frac{4x}{\pi} \int_0^{\sqrt{2-x}} du \sqrt{\frac{(u^2 + x)(2 - x - u^2)}{u^2 + 2x}}. \quad (21)$$

With the above assumptions on the orientations of the orbital planes and the distribution of eccentricities it can be shown that the *average* observed projected separation  $\langle s \rangle$  for a given semi-major axis  $a$  is basically equal to  $a$  (e.g. Yoo et al. 2004):

$$\langle s \rangle = \frac{5\pi}{16} a \simeq 0.98a. \quad (22)$$

<sup>12</sup> We use the term “solar neighbourhood” for local quantities.



**Fig. 4.** Effective volumes calculated using the parameters that correspond to our final sample (Table 2) for the three Galactic structure parameter sets as discussed in §4.2. Long dashed line: set 1; solid line: set 2; short dashed line: set 3.

From the normalisations of  $q(a)$  it also follows that  $Q(s)$  is normalised to unity in the range  $[\langle s \rangle_{\min}, \langle s \rangle_{\max}]$ .

Substituting  $\eta = \sqrt{s/2a}$  in Eq. 20, as well as  $\xi = u/\sqrt{2}$  in Eq. 21 leads to a convenient form for numerical integration

$$Q(s) = \left(\frac{s}{\text{pc}}\right)^{-\lambda} C_\lambda(s) \text{ pc}^{-1}, \quad (23)$$

where

$$C_\lambda(s) = 2^{\lambda+5} \frac{c_\lambda}{\pi} \int_0^{\min(1, \sqrt{s/2a_{\min}})} d\eta \eta^{2\lambda+1} \int_0^{\sqrt{1-\eta^2}} d\xi \sqrt{\frac{(\xi^2 + \eta^2)(1 - \eta^2 - \xi^2)}{\xi^2 + 2\eta^2}}. \quad (24)$$

The integrals in Eq. 24 are evaluated using Gauss quadrature (Press et al. 1992). For  $s$  far from  $a_{\min}$  or  $a_T$  the distribution of the projected separations  $Q(s)$  is approximately powerlaw

$$Q(s) \propto \left(\frac{s}{\text{pc}}\right)^{-\lambda} \text{ pc}^{-1}. \quad (25)$$

#### 4.1.2. Effective volume

As we deal with a magnitude-limited sample, we are plagued by selection effects that must be properly taken into account. Following WW87, we do this by means of the *effective volume*. It allows for the solid angle  $\Omega$  covered by the sample, the angular separation range  $[\theta_{\min}, \theta_{\max}]$  we are examining, and the apparent magnitude limits  $m_{\min}$  and  $m_{\max}$ . Furthermore, the effective volume takes the stellar density distribution  $\rho$  into account, as well as the normalised (single star) luminosity function  $\tilde{\Phi}$  (both described more in detail in Sect. 4.2).

Making the plausible assumptions that the intrinsic luminosities of the two stars in a wide binary are independent and distributed in the same way as the luminosities of field stars

(WW87), and assuming further that the stellar density distribution  $\rho$  is independent of the stars' absolute luminosities, we may write the effective volume as

$$V(s) = \Omega \int_{s/\theta_{\min}}^{s/\theta_{\max}} dD D^2 \tilde{\rho}(D) \iint_{M_{\min}(D)}^{M_{\max}(D)} dM_1 dM_2 \tilde{\Phi}(M_1) \tilde{\Phi}(M_2), \quad (26)$$

where we have the standard relations

$$M_{\max}(D) = m_{\max} - 5 \log_{10}(D/10 \text{ pc}) \quad (27)$$

$$M_{\min}(D) = m_{\min} - 5 \log_{10}(D/10 \text{ pc}), \quad (28)$$

and  $\tilde{\rho}(D)$  is normalised such that  $\tilde{\rho}(0) = 1$ .

In Fig. 4 the effective volumes for the three Galactic structure parameter sets (described in the next section) are plotted using the parameters that correspond to our final sample (see Table 2). From the effective volume  $V(s)$ , it is evident that our study is most sensitive to projected separations in  $0.01 \text{ pc} \lesssim s \lesssim 0.1 \text{ pc}$ , whereas our study is absolutely insensitive when  $s \lesssim 0.001 \text{ pc}$  or  $s \gtrsim 1 \text{ pc}$ . Referring to Eq. 22, we see that the shape of the effective volume also indicates that we are insensitive to semi-major axis  $a \lesssim 0.001 \text{ pc}$  and for  $a \gtrsim 1 \text{ pc}$ , which nicely fits the range in  $a$  where we assume that the single power-law model holds. But it also shows that our analysis is not too sensitive to very wide binary stars with semi-major axes larger than  $0.1 \text{ pc}$ .

## 4.2. Galactic model

### 4.2.1. Stellar density distribution

Following earlier work (especially Jurić et al. 2008; Chen et al. 2001, and references therein), we modelled the stellar density distribution of the Milky Way Galaxy by including two exponential disc components – a thin and a thick disc – and an elliptical halo component whose density profile obeys a powerlaw. The contribution of the Galactic bulge is negligible in the direction of the NGP and it is therefore ignored in the following.

The three components are added to yield the total density distribution

$$\rho(D) = \rho(0) \tilde{\rho}(D) \quad (29)$$

with

$$\tilde{\rho}(D) = \frac{\tilde{\rho}^{\text{thin}}(D) + n^{\text{thick}} \tilde{\rho}^{\text{thick}}(D) + n^{\text{halo}} \tilde{\rho}^{\text{halo}}(D)}{1 + n^{\text{thick}} + n^{\text{halo}}}, \quad (30)$$

which is normalised such that  $\tilde{\rho}(0) = 1$ . The thick disc and the halo component are normalised with respect to the thin disc via the normalisation constants  $n^{\text{thick}}$  and  $n^{\text{halo}}$ , respectively. The overall normalisation  $\rho(0)$  is calibrated to produce the observed star counts.

Both the thin and the thick disc populations obey a double exponential density law of the form (e.g. Bahcall & Soneira 1980)

$$\exp\left(-\frac{r - r_0}{h_r}\right) \exp\left(-\frac{|z|}{h_z}\right) \quad (31)$$

where  $h_r$  and  $h_z$  are the scale length and the scale height of the disc,  $z$  is the object's height above the Galactic midplane, and  $r$  is its Galactocentric distance in the Galactic plane. We take the Sun's distance from the centre of the Galaxy in the Galactic plane to be  $r_0 = 8 \text{ kpc}$ . Jurić et al. (2008) find that the Sun is located  $z_0 \approx 25 \text{ pc}$  above the Galactic midplane (Chen et al. (2001)

give  $z_0 \simeq 27$  pc). It is straightforward to derive the following useful relations

$$z = z_0 + D \sin b \quad (32)$$

$$r^2 = r_0^2 + d^2 - 2r_0d \cos \ell \quad (33)$$

$$d = D \cos b, \quad (34)$$

where  $d$  is its distance from the Sun in the Galactic plane and  $\ell$  and  $b$  are its Galactic longitude and latitude, respectively. The stellar density is fairly constant within the sample, so we neglect the direction-dependent density variations. For the sake of simplicity, we adopt the coordinates of our sample's centre:  $\ell = 175.6^\circ$  and  $b = 81.6^\circ$ . (Dividing our sample into subsamples and summing over them taking their centres hardly influences our results.)

In line with Jurić et al. (2008) and Chen et al. (2001) we assume the scale heights to be independent of absolute magnitude  $M$ . Using capitals for the thick disc's and lower case letters for the thin disc's scale height and length, we may write the normalised number density distribution of the stars in the Galactic thin disc as

$$\tilde{\rho}^{\text{thin}}(D) = \exp\left(-\frac{r - r_0}{h_r} - \frac{|z|}{h_z}\right) \cdot \exp\left(\frac{z_0}{h_z}\right) \quad (35)$$

and that of the thick disc as

$$\tilde{\rho}^{\text{thick}}(D) = \exp\left(-\frac{r - r_0}{H_r} - \frac{|z|}{H_z}\right) \cdot \exp\left(\frac{z_0}{H_z}\right), \quad (36)$$

where the rightmost factors are for normalisation purposes and ensure that  $\tilde{\rho}^{\text{thin}}(0) = \tilde{\rho}^{\text{thick}}(0) = 1$ .

The density distribution of the stellar halo population is modelled by a powerlaw with index  $k$ . The observational data prefer a somewhat oblate halo giving an ellipsoid, flattened in the same sense as the Galactic disc, with axes  $a = b$  and  $c = \kappa a$ , where  $\kappa$  controls the ellipticity of the halo (cf. Jurić et al. 2008)

$$\tilde{\rho}^{\text{halo}}(D) = \left[ \frac{r^2 + \left(\frac{z}{\kappa}\right)^2}{r_0^2 + \left(\frac{z_0}{\kappa}\right)^2} \right]^{-\frac{k}{2}}, \quad (37)$$

which, of course, also satisfies  $\tilde{\rho}^{\text{halo}}(0) = 1$ .

We compare three different sets of structure parameters:

*Set 1:* The *measured* values from Jurić et al. (2008) (see their Table 10). These values are best-fit parameters as *directly* measured from the apparent number density distribution maps using their “bright” photometric parallax relation (their Eq. 2). They are not corrected for bias caused by, e.g., unresolved stellar multiplicity, hence the term “apparent”.

*Set 2:* The *bias-corrected* values from Jurić et al. (2008) (see again their Table 10). These values were corrected for unrecognised stellar multiplicity, Malmquist bias, and systematic distance determination errors by means of Monte Carlo-generated mock catalogues (Jurić et al. 2008, §4.3.). Following Reid & Gizis (1997), the fraction of “stars” in the local stellar population that in fact are unresolved binaries is taken to be 35%. However, the halo component was not included in the mock catalogues, and its structure parameters were therefore not corrected, but the measured values are used instead. This set of parameters will be referred to as our standard set.

*Set 3:* As a third independent set of Galactic structure parameters, we refer to the somewhat earlier work of Chen et al. (2001). It is also based on observations obtained with the SDSS, but Chen et al. (2001) infer the density distribution of the stars

**Table 1.** Galactic structure parameters

structure parameter	set 1	set 2	set 3
$z_0$ [pc]	25	25	27
$h_z$ [pc]	245	300	330
$h_r$ [pc]	2 159	2 600	2 250
$n^{\text{thick}}$	0.13	0.12	0.0975
$H_z$ [pc]	743	900	665
$H_r$ [pc]	3 261	3 600	3 500
$n^{\text{halo}}$	0.0051	0.0051	0.00125
$\kappa$	0.64	0.64	0.55
$k$	2.77	2.77	2.5

by inverting the fundamental equation of stellar statistics (e.g. Karttunen et al. 1996).

These three sets of Galactic structure parameters are summarised in Table 1. Jurić et al. (2008) quote unrecognised multiplicity as one of the dominant sources of error in the distance determination of the stars. (Only the uncertainties in absolute calibration of the photometric parallax relation might be even more important, but little can be done to increase its accuracy at the moment; see, however, Sesar et al. (2008).) It is therefore not surprising that the scale heights and lengths of the bias-corrected parameter set are larger than those of the measured parameters as the misidentification of a binary star as a single star results in an underestimation of its distance. Regarding the values derived by Chen et al. (2001), we find the largest difference in the normalisation of the halo component that is about four times smaller than that given by Jurić et al. (2008), whereas the other parameters are not too dissimilar.

#### 4.2.2. Stellar luminosity function

For our study we refer to the luminosity function (LF) inferred by Jahreiß & Wielen (1997) using Hipparcos parallaxes, which gives reliable values for a wide magnitude range ( $-1 \leq M_V \leq 19$ ). (The faint end of the LF is somewhat uncertain and Jahreiß & Wielen (1997) give only a lower limit in the range  $20 \leq M_V \leq 23$ . We take that lower limit at the faint end to be the true value of the LF in that magnitude range.)

We need to transform the Jahreiß & Wielen (1997) LF from visual ( $V$ -band) into  $r$ -band magnitudes, i.e. from  $\Phi(M_V)$  to  $\Phi(M_r)$ . We perform this transformation in an unsophisticated way by combining the photometric parallax relation from Laird et al. (1988), which was calibrated using the Hyades and is linear in  $(B - V)$  (their Eq. 1a)

$$M_V = 5.64(B - V) + 1.11 \text{ mag} \quad (38)$$

with a transformation equation that is obtained by subtracting Eq. 1 from 3 in Bilir et al. (2005)

$$V - r = 0.491(B - V) - 0.144 \text{ mag}. \quad (39)$$

The linearity of the photometric parallax relation removes any inversion problem, since it assures – in a mathematical sense – that a colour index  $(B - V)$  exists for every  $M_V$ . For a given distance we then have

$$M_r = 0.913M_V + 0.0476 \text{ mag} \quad (40)$$

and the transformed LF is simply

$$\Phi(M_r) = \Phi(M_V) \frac{dM_V}{dM_r}. \quad (41)$$

Using linear interpolation (the brightest bin at  $M_r = -1$  mag is an *extrapolation*), we finally have rebinned  $\Phi(M_r)$  such that the bins are centred on integer values of absolute magnitude  $M_r$  and have a width of  $\Delta M_r = 1$  mag.

We normalise the LF by integrating over the absolute magnitude range corresponding to our total sample

$$\tilde{\Phi}(M_r) = \frac{\Phi(M_r)}{\int_{5.2}^{\infty} \Phi(\mu) d\mu} = \frac{\Phi(M_r)}{n'_*} \quad (42)$$

with  $n'_* \simeq 0.094 \text{ pc}^{-3}$  being the total number density of stars in the solar neighbourhood having  $g - r > 0.5$  mag. Assuming that all stars are on the MS, we find, using the (“bright”) photometric parallax relation from Jurić et al. (2008), that the cut at  $g - r = 0.5$  mag corresponds to  $M_r \simeq 5.2$  mag. (We first transformed it from  $g - r$  into  $r - i$  using Eq. 4 from Jurić et al.)

#### 4.3. Modification of the Wasserman-Weinberg technique

To derive a theoretical angular 2PCF we need to calculate the expected number of wide binaries as a function of *angular* separation. To this end, the above-described WW-technique requires the modification that we address now.

Let  $\varphi(\theta)d\theta$  be the number of wide binaries observed with an angular separation between  $\theta$  and  $\theta + d\theta$ . For a given distance  $D$ , we have  $\varphi(\theta)d\theta = \psi(s)ds$  and  $s = D\theta$ . (Note that our unit for  $s$  and  $D$  is pc, whereas  $\theta$  is in rad.) The latter expression can be used to write the reduced separation distribution  $Q(s)$  as  $Q(D\theta)$ . This introduces an explicit dependency on  $D$ , so we need to incorporate  $Q(D\theta)$  into the integration over  $D$  in the formula for the effective volume (Eq. 26). This incorporation is the reason it is no longer possible to separate the reduced separation distribution from the effective volume in a formal way like in Eq. 15.

Furthermore, we need to modify the limits in the integration over  $D$  in Eq. 15. Recalling that for a given semi-major axis  $a$ , the average observed projected separation is  $\langle s \rangle = 0.98a$  (Eq. 22), it appears to be appropriate to let the integration limits run from  $\langle s \rangle_{\min}/\theta$  to  $\langle s \rangle_{\max}/\theta$  (Garnavich 1991). We take  $\langle s \rangle_{\min} = 0.98a_{\min} = 9.8 \cdot 10^{-4}$  pc and  $\langle s \rangle_{\max} = 0.98a_{\mathrm{T}}$  (for the calculation of the tidal limit  $a_{\mathrm{T}}$  see Appendix B).

Putting it all together, we may write the number of observed wide binaries as a function of angular separation as

$$\varphi(\theta) = n_{\mathrm{WB}} \Omega \int_{\langle s \rangle_{\min}/\theta}^{\langle s \rangle_{\max}/\theta} dD D^3 \tilde{\rho}(D) Q(D\theta) \iint_{M_{\min}(D)}^{M_{\max}(D)} dM_1 dM_2 \tilde{\Phi}(M_1) \tilde{\Phi}(M_2), \quad (43)$$

where we have included an additional factor  $D$  into the integration over  $D$ , because  $ds = Dd\theta$  for a given  $D$ .

The model-2PCF is now determined by adding the number of physical pairs, calculated by Eq. 43, to the number of pairs expected from a random sample given by Eq. 6

$$w_{\mathrm{mod}}(\theta) = \frac{\varphi(\theta) + P(\theta)}{P(\theta)} - 1 = \frac{\Omega \varphi(\theta)}{\pi N_{\mathrm{obs}}(N_{\mathrm{obs}} - 1)\theta}. \quad (44)$$

The two free parameters –  $n_{\mathrm{WB}}$  and  $\lambda$  – are determined in a least-square sense by fitting the measured 2PCF  $\hat{w}_{\mathrm{corr}}(\theta)$  to the model 2PCF  $w_{\mathrm{mod}}(\theta)$  as described more in detail in the next section.

#### 4.4. Fitting procedure

We determine the two free parameters of the model,  $n_{\mathrm{WB}}$  and  $\lambda$ , by means of a Levenberg-Marquardt nonlinear least-square

**Table 3.** Best-fit values: total sample

set of structure parameters	$n_{\mathrm{WB}}$ [pc $^{-3}$ ]	$\lambda$	$Q$	$\chi^2/\nu$	$f_{\mathrm{WB}}\%$
set 1	0.0052	1.00	0.29	1.14	10.2
set 2	0.0069	1.01	0.30	1.13	13.6
set 3	0.0061	1.01	0.29	1.14	12.1

algorithm (Lourakis 2004)<sup>13</sup>, which minimises the value of the  $\chi^2$  defined as

$$\chi^2 = \sum_{i=1}^N \left( \frac{\hat{w}_{\mathrm{corr}}(\theta_i) - w_{\mathrm{mod}}(\theta_i)}{\delta \hat{w}_{\mathrm{corr}}(\theta_i)} \right)^2. \quad (45)$$

The data is binned in steps of  $\Delta\theta \equiv \theta_{i+1} - \theta_i = 1''$ , where  $\theta_1 = \theta_{\min} = 2''$  and  $\theta_N = \theta_{\max} = 30''$  as imposed by the resolution limit of the SDSS and maximal distance in the *Neighbors* table, respectively. As a result, we have  $N = 28$  here. Given the approximative character of our study, the use of the this standard definition of the  $\chi^2$  is appropriate, even though the values of the 2PCF at different angular separations are not strictly independent of each other.

Following Press et al. (1992), we use the incomplete gamma function  $Q(\chi^2|\nu)$  with  $\nu = N - 2 = 26$  degrees of freedom as a quantitative measure of the goodness-of-fit. ( $N-2$  because the model has two free parameters:  $n_{\mathrm{WB}}$  and  $\lambda$ .) Values of  $Q$  near unity indicate that the model adequately represents the data.

#### 4.5. Confidence intervals

To estimate the uncertainties of our best-fit values, we use Monte Carlo confidence intervals (MCCRs) (e.g. Press et al. 1992, §15.6). Assuming Poissonian errors, we generate 10 000 synthetic data sets by drawing the number of unique pairs “observed” in the  $k$ -th synthetic sample  $F_{\mathrm{syn}}^{(k)}(\theta)$  from a Poisson distribution with mean  $F(\theta)$ , where  $F(\theta)$  is the observed number of pairs, *not* corrected for edge effects due to survey holes. The “true” number of pairs in a synthetic sample is then given by dividing  $F_{\mathrm{syn}}^{(k)}$  by  $F_{\mathrm{H}}^{\mathrm{tot}}$  (see §3.2).

For each synthetic sample, we determine best-fit values for the model’s free parameters,  $n_{\mathrm{WB}}^{(k)}$  and  $\lambda^{(k)}$ , in a least-square sense as described above. A  $p\%$ -MCCR is defined by the line of constant  $\chi^2$  which encloses  $p\%$  of the best-fit values in the  $n_{\mathrm{WB}}$  versus  $\lambda$  plane (see Fig. 6). The confidence intervals of  $n_{\mathrm{WB}}$  and  $\lambda$  are then given by the orthogonal projection of the MCCR onto the corresponding axis.

We include in our error estimate only the uncertainties stemming from the pair counts in  $F(\theta)$ . Neither the uncertainties in the Galactic structure parameters nor those in the LF are taken into account.

## 5. Results

### 5.1. Analysis of the total sample

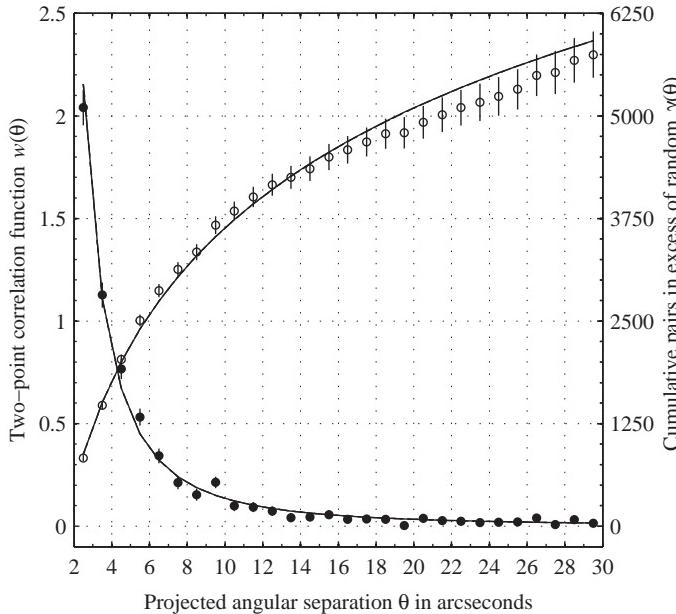
The results of the analysis for the total sample are shown in Fig. 5 and Table 3. Figure 5 shows the 2PCF estimate  $\hat{w}_{\mathrm{corr}}(\theta)$  and the CDD  $\hat{\gamma}(\theta)$ . The statistical uncertainties calculated according to Eq. 13 are shown as vertical lines. A strong clustering signal out to at least  $\theta = 10''$  is evident, whereas the CDD suggest that there are pairs in excess of a random distribution up to

<sup>13</sup> We use *levmar-2.2*.

**Table 2.** Parameters of the final sample and of the subsamples

(Sub)sample	$\Omega$ [deg $^2$ ]	$\ell$ [deg]	$b$ [deg]	$N_{\text{obs}}$	$D_{\text{eff}}^{\text{a}}$ [pc]	$a_T^{\text{a}}$ [pc]	$f_{\text{Halo}}^{\text{a}}$ %
total	668.58	175.6	81.6	669 843	1 555	1.07	30.0
$r < 20.0$ mag	668.58	175.6	81.6	535 595	1 380	1.06	25.2
$r < 19.5$ mag	668.58	175.6	81.6	425 674	1 235	1.06	20.6
left	334.57	191.2	73.7	326 333	1 495	1.08	28.3
right	334.01	102.5	84.8	343 510	1 625	1.05	31.0
A	78.94	180.1	68.3	77 008	1 460	1.09	26.8
B	78.10	167.5	75.3	76 011	1 510	1.08	28.6
C	78.99	136.0	79.8	76 686	1 565	1.06	29.9
D	78.81	93.3	78.4	82 780	1 635	1.05	30.6
E	88.23	208.0	69.7	87 751	1 485	1.08	27.4
F	88.30	210.6	78.6	85 563	1 545	1.07	29.5
G	87.95	215.2	87.5	88 440	1 615	1.05	31.0
H	88.27	33.4	83.6	95 604	1 700	1.04	31.8

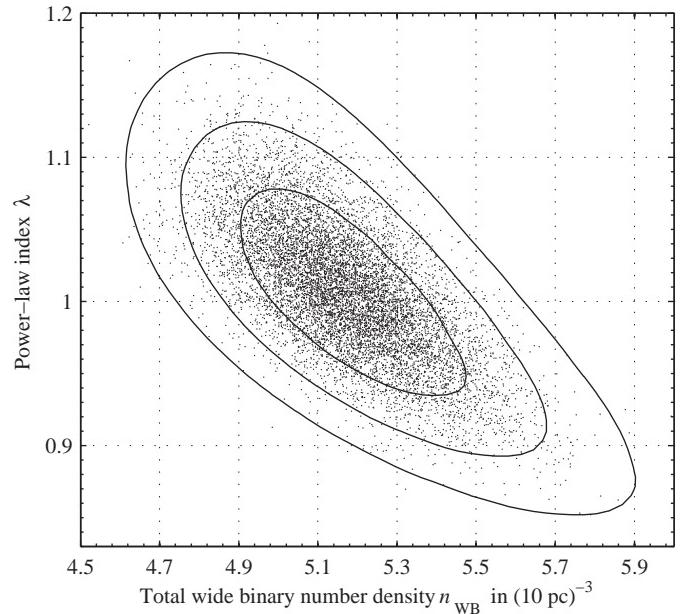
<sup>a</sup> Calculated using the Galactic structure parameter set 2.



**Fig. 5.** 2PCF as inferred from the total sample (solid circles, left ordinate) and the corresponding CDD (open circles, right ordinate). Poisson errors are indicated as vertical lines. Model curves for the three Galactic structure parameter sets are plotted, too, but as the differences between them are marginal they lie one upon the other, giving a single solid line.

maximum angular separation examined, that is, up to 30''. The outlier at  $\theta = 9''$  is probably a random fluctuation. We also plot in Fig. 5 best-fitting models using the three Galactic structure parameter sets described in Sect. 4.2.1.

The best-fit values of the two free parameters,  $n_{\text{WB}}$  and  $\lambda$ , are tabulated in Table 3 for the three Galactic structure parameter sets. The power-law index  $\lambda$  appears to be quite independent of the set we choose. The wide binary density  $n_{\text{WB}}$ , on the other hand, shows some variation. The difference between the sets 1 and 2 reflects, for the most part, the difference in the overall normalisation  $\rho(0)$  of the density distribution, whereas in set 3 the unequal halo normalisation with respect to the other two sets also contributes to the difference in  $n_{\text{WB}}$ .



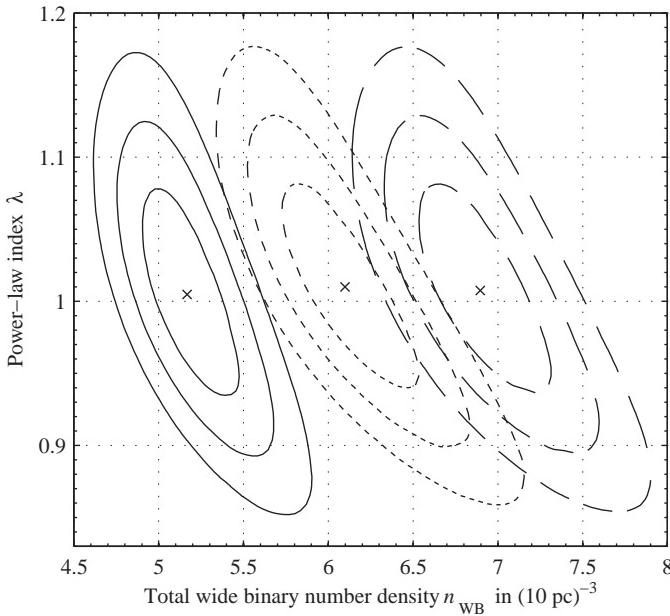
**Fig. 6.** Distribution of the best-fit values for the total sample using the structure parameter set 2 obtained by the Monte Carlo procedure described in the text. The solid contours (MCCRs) are lines of constant  $\chi^2$  and enclose 68.3%, 95.4%, and 99.7% of the best-fit values.

We also list the goodness-of-fit  $Q$  and the corresponding reduced chi-square values ( $\chi^2$  divided by the degrees of freedom  $v$ ) in Table 3. All three sets of Galactic structure parameters give equally good fits, whereas the reduced chi-square values, which are only slightly more than unity, indicate that we have not severely underestimated the uncertainty in the 2PCF.

In Fig. 6 we show, representative of the other structure parameter sets, the distribution of the best-fit values from the synthetic samples using set 2, our standard set. In the same figure the 68.3% ( $1\sigma$ ), 95.4% ( $2\sigma$ ), and 99.7% ( $3\sigma$ ) MCCRs are also plotted. Quoting 95.4% confidence intervals throughout, we find for our final sample using the Galactic structure parameter set 2

$$n_{\text{WB}} = 0.0052^{+0.0006}_{-0.0005} \text{ pc}^{-3} \quad \text{and} \quad \lambda = 1.00^{+0.13}_{-0.12}. \quad (46)$$

The power-law index  $\lambda$  of semi-major axis distribution is consistent with Öpik's law ( $\lambda = 1$ ) up to the Galactic tidal limit,



**Fig. 7.** MCCR for the three Galactic structure parameter sets corresponding to the total sample. Long dashed line: set 1; solid line: set 2; short dashed line: set 3. The crosses indicate the best-fit values inferred from the true *observed* number of pairs,  $F_{\text{corr}}(\theta)$ .

whereas the number density  $n_{\text{WB}}$  corresponds to a wide binary fraction with respect to *all* stars (*no* colour-cut) in the solar neighbourhood of

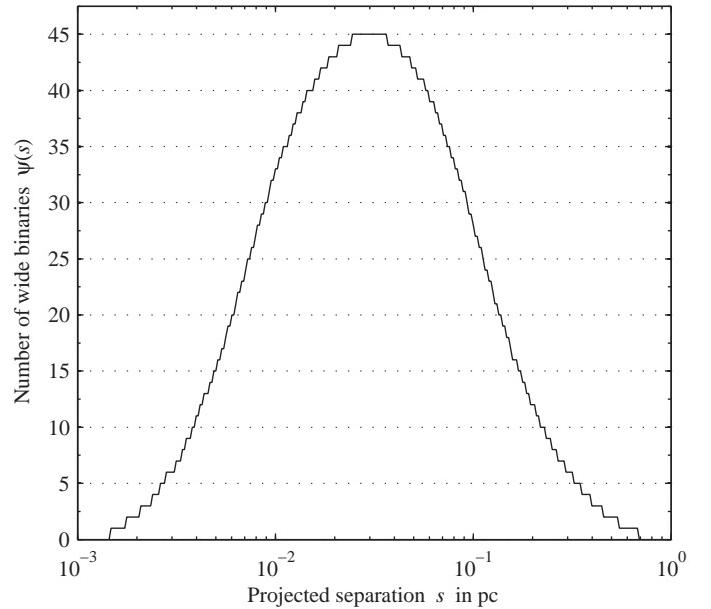
$$f_{\text{WB}} \equiv \frac{2n_{\text{WB}}}{n_*} = 10.2^{+1.2}_{-1.0}\%, \quad (47)$$

where  $n_* \simeq 0.10 \text{ pc}^{-3}$  is the *total* local stellar number density, i.e. the integral over the *whole* Jahreiß & Wielen LF. (The local wide binary density  $n_{\text{WB}}$  corresponds only to wide binaries with both components having  $g - r > 0.5$  mag.)

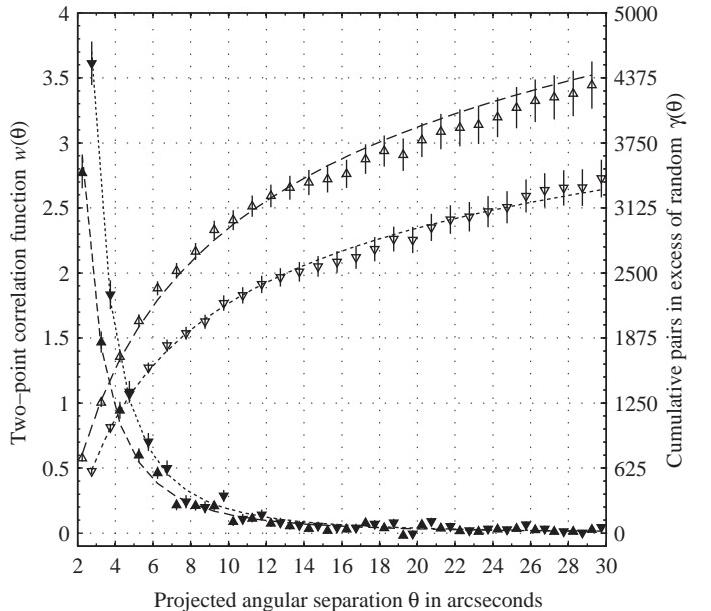
We show the confidence regions corresponding to the three sets of Galactic structure parameters in Fig. 7. The confidence intervals for  $\lambda$  agree for all the sets, which demonstrates that we can determine  $\lambda$  and its confidence intervals reliably – provided that no systematic error has crept into our analysis. The wide binary density  $n_{\text{WB}}$ , on the other hand, is more sensitive to the exact values of the Galactic structure parameters than  $\lambda$  is, so the values we have derived for the wide binary density  $n_{\text{WB}}$  and the fraction  $f_{\text{WB}}$  should be viewed with caution. We are, however, confident that the true value of the wide binary density  $n_{\text{WB}}$  is within a factor of 2 of our derived value.

Since the structure parameter set 2 is – as far as known to the authors – the only one systematically corrected for unresolved multiplicity, we give that set more weight. From now on, we use the bias-corrected values from Jurić et al. (2008), i.e. our standard set 2, exclusively.

Having determined the two free parameter in our model, we can calculate the number of wide binaries in our total sample as a function of the projected separation  $s$ . We show the distribution in Fig. 8 whose shape is largely dominated by the effective volume, i.e. by selection effects. In particular, this distribution implies that we have observed  $830^{+250}_{-215}$  very wide binaries with a projected separation larger than 0.1 pc in our total sample, while none are expected to be found beyond 0.8 pc, given our selection criteria. However, that extremely wide binaries with projected



**Fig. 8.** Distribution of wide binaries as a function of projected separation  $s$  that is expected to be observed in the total sample, given the model assumptions and the selection criteria we have adopted (see text).

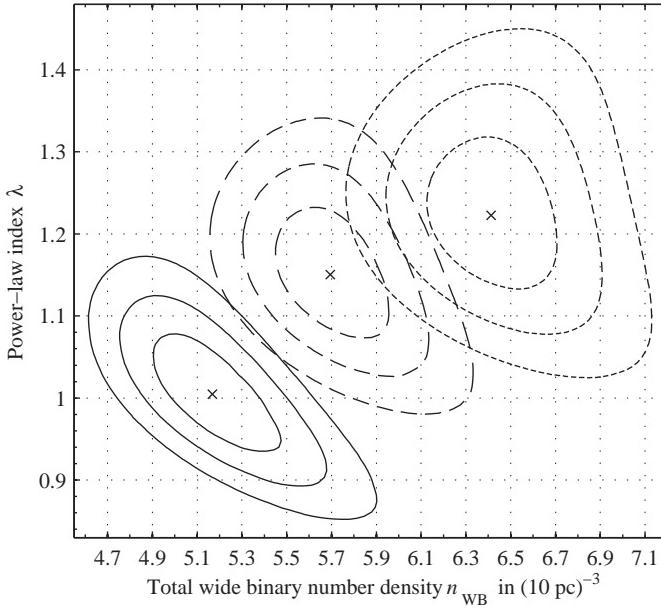


**Fig. 9.** 2PCF and CDD for the  $r < 20.0$  mag and  $r < 19.5$  mag subsamples together with the corresponding model curves. Up-pointing triangles (long-dashed lines) are for the  $r < 20.0$  mag subsample, whereas down-pointing triangles (short-dashed lines) are for the  $r < 19.5$  mag subsample. The symbols were shifted apart by 0.25'' for better visibility.

separations of more than 1 pc can exist in the Galactic halo has recently been confirmed by Quinn et al. (2009).

## 5.2. Differentiation in terms of apparent magnitude

To test the consistency of the standard model, we used two subsamples having brighter upper apparent magnitude limits than



**Fig. 10.** MCCR for the  $r < 20.0$  mag (long dashed contours) and  $r < 19.5$  mag (short-dashed contours) subsamples. For comparison, the MCCR for our final sample is also shown (solid contours). The crosses have the same meaning as in Fig. 7.

**Table 4.** Best-fit values:  $r < 20.0$  mag and  $r < 19.5$  mag

Subsample	$n_{WB}$ [pc <sup>-3</sup> ]	$\lambda$	$Q$	$\chi^2/\nu$	$f_{WB}$ %
$r < 20.0$ mag	0.0057	1.15	0.18	1.25	11.3
$r < 19.5$ mag	0.0064	1.22	0.34	1.09	12.7

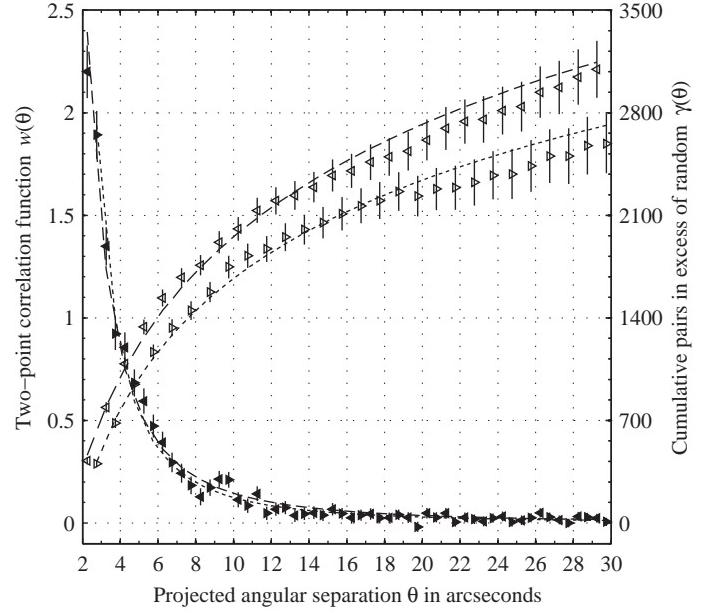
**Table 5.** Best-fit values: left and right

Subsample	$n_{WB}$ [pc <sup>-3</sup> ]	$\lambda$	$Q$	$\chi^2/\nu$	$f_{WB}$ %
left	0.0055	1.00	0.55	0.94	10.8
right	0.0046	1.03	0.82	0.74	9.1

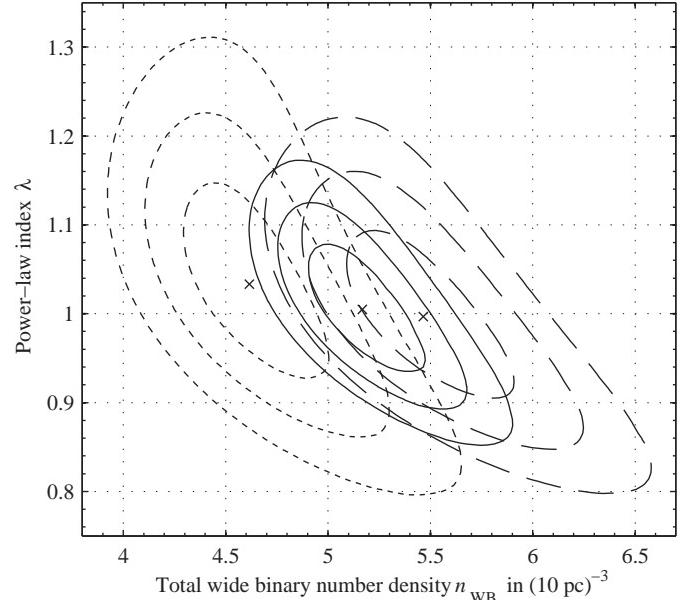
our total sample and analysed them in the same way as the total sample itself. We set the upper apparent magnitude limit to  $r = 20.0$  mag and  $r = 19.5$  mag – all other selection criteria (including the lower apparent magnitude limits) remain unchanged. The main parameters of these subsamples are listed in Table 2.

In Fig. 9 we show the 2PCF and the CDD as in Fig. 5, together with the corresponding model curves for standard set 2 (see caption). The best-fit values are listed in Table 4. The MCCR of the  $r < 20.0$  mag and  $r < 19.5$  mag subsamples are shown in Fig. 10.

If the model were self-consistent, we would expect that the best-fit values agree with each other within their uncertainties. Figure 10 indicates that this is not the case: It appears that the best-fit values are systematically shifted to higher densities and larger power-law indices when using a brighter upper magnitude limit. As we discuss in Sect. 6.2, this inconsistency is most likely an artefact of the (oversimplifying) model assumptions and does not entirely undermine our results.



**Fig. 11.** 2PCF and CDD for the left and the right subsamples, together with the corresponding model curves. Left-pointing triangles (long-dashed lines) are for the left subsample, whereas right-pointing triangles (short-dashed lines) are for the right subsample. The symbols were shifted apart by 0.25'' for better visibility.



**Fig. 12.** MCCR for the left (long-dashed contours) and right (short-dashed contours) subsamples. For comparison, also the MCCR for our final sample is shown (solid contours). The crosses have the same meaning as in Fig. 7.

### 5.3. Differentiation in terms of direction

In a previous study, Saarinen & Gilmore (1989) found that the binaries appear to be highly clumped in the NGP. However, it has not become entirely clear whether this patchiness of the wide binary distribution in the sky is a real physical characteristic of the wide binary population or if it is due to statistical fluctuations. In

**Table 6.** Best-fit values: A to H

Subsample	$n_{WB}$ [pc $^{-3}$ ]	$\lambda$	$Q$	$\chi^2/v$	$f_{WB}$ %
A	0.0050	1.27	0.92	0.64	9.9
B	0.0054	1.00	0.96	0.57	10.7
C	0.0041	1.02	0.36	1.07	8.2
D	0.0046	1.19	0.29	1.13	9.1
E	0.0077	0.73	0.24	1.18	15.2
F	0.0045	1.14	0.37	1.07	8.9
G	0.0054	0.83	0.67	0.86	10.6
H	0.0042	1.27	0.64	0.88	8.4

principle, we can check whether the wide binary density varies with position in the sky by dividing our sample into subareas.

To begin with, we divide our sample in two halves by cutting it along the  $\alpha = 185^\circ$  meridian. In the following, we refer to the subarea with  $\alpha < 185^\circ$  as the “left” subsample and the other half as the “right” subsample. The subsamples’ main parameters are listed in Table 2.

In Fig. 11 we show the 2PCFs inferred from the left and the right sample, as well as the corresponding CDDs. It appears that there are a few more pairs in excess of random in the left half. In the same figure also the corresponding model curves are plotted. The best-fit values are listed in Table 5.

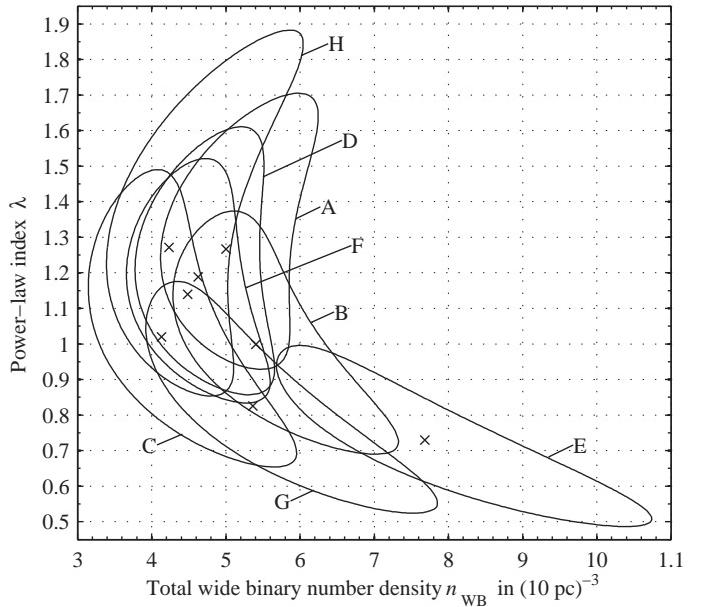
In Fig. 12 we show the MCCR of the left and the right subsamples. The left subsample indeed shows a higher wide binary density than the right one. The difference is significant at the  $3\sigma$  level. The power-law indices, on the other hand, do agree in both subsamples. Regarding the total sample, the right half differs significantly (at  $3\sigma$ ), whereas the left half is more consistent with it.

The difference in the wide binary density between the left and the right halves is probably a real feature, as any inadequacies of our model (e.g. inaccuracies of the stellar density distribution we use) should affect both halves in almost the same manner. To examine this apparent positional dependency of the wide binary density in more detail, we divide our sample further into eight subsamples, each covering  $10^\circ \times 10^\circ$ . They are labelled from A (“upper left”) to H (“lower right”) as suggested in Fig. 13, where the abscissa can thought of representing the right ascension from  $165^\circ$  (“left”) to  $205^\circ$  (“right”) and similarly the ordinate represents the declination from  $22^\circ$  (“bottom”) to  $42^\circ$  (“top”). The subsample’s main parameters are again summarised in Table 2.

Figure 13 shows the 2PCFs estimate and the CDDs with the best-fit model curves. The corresponding best-fit parameters are listed in Table 6. Some subtle differences are apparent between different subsamples. While the CDD in A, B, C, and G are reproduced well by the model, the CDD in D, F, and H appears to be too flat. In those subsamples it seems that almost no physical pairs are present at angular separations over  $15''$ , in agreement with the findings of Sesar et al. (2008) (see Sect. 6.1).

The subsample E appears to be an outlier, because it contains almost twice as many pairs in excess of random as the seven other subsamples. The listed best-fit values confirm the peculiar character of subsample E having the highest wide binary density of all eight subsamples, together with the lowest power-law index. As far as the authors can judge, there is no obvious special feature (e.g. open star cluster) in subsample E that could cause this anomaly.

The  $2\sigma$  MCCR are shown in Fig. 14. Except for the outlier E, all the other subsamples are quite consistent with each other. Also, no obvious trend, e.g. with Galactic latitude  $b$ , is apparent.

**Fig. 14.** 95.4% MCCR for the subsamples A to H. The crosses have the same meaning as in Fig. 7.**Table 7.** Best-fit values: excluding subsample E

Subsample	$n_{WB}$ [pc $^{-3}$ ]	$\lambda$	$Q$	$\chi^2/v$	$f_{WB}$ %
total\ E	0.0049	1.06	0.40	1.04	9.6
left\ E	0.0049	1.12	0.76	0.79	9.8

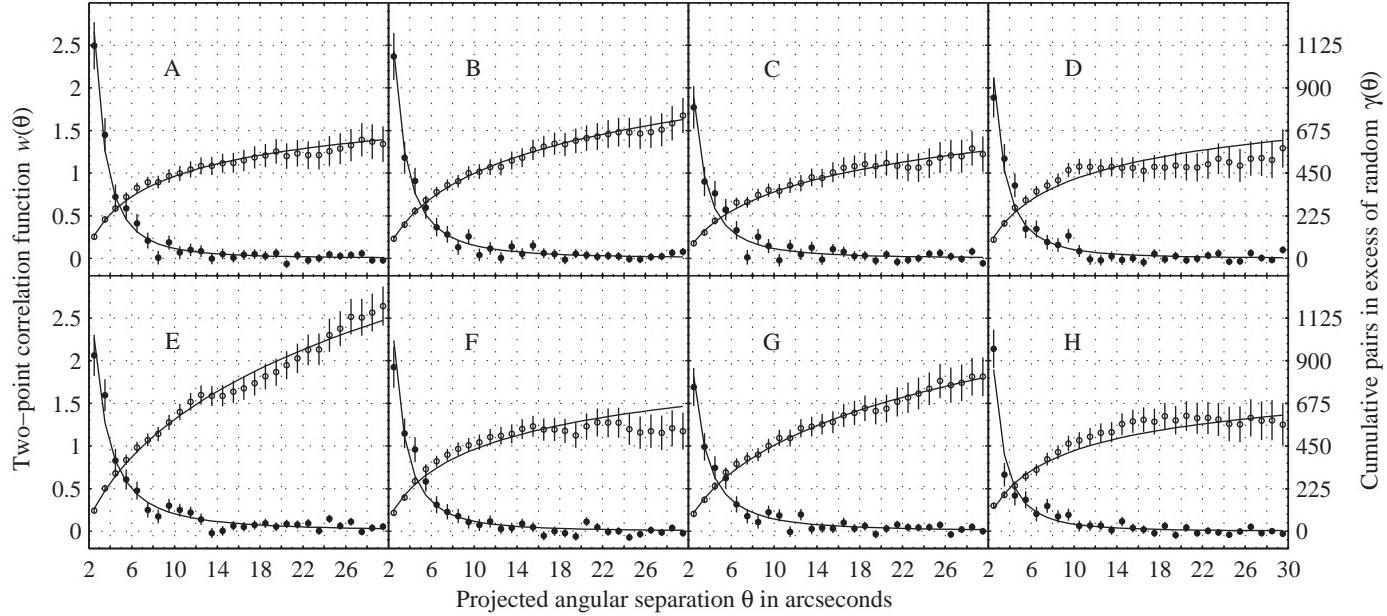
To what extent is the subsample E responsible for the difference between the right and the left subsamples?

To answer this question, we repeated the analysis of the total and left (sub)sample excluding subsample E. The results are listed in Table 7. The wide binary density drops significantly to a value more consistent with the right subsample. Thus, we conclude that the difference between the right and the left subsamples is largely caused by the outlier E. Apart from subsample E, the wide binary densities in different directions appear to be consistent with each other. The reason for the high density in subsample E remains unclear. However, a statistical fluctuation cannot be ruled out to a level better than  $2\sigma$ .

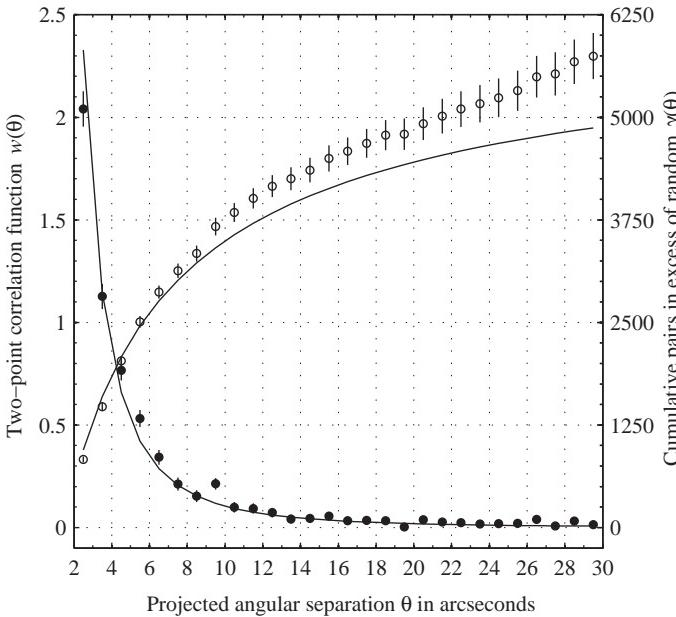
## 6. Discussion

### 6.1. External (in)consistencies

In this section we compare our results with those of previous studies. Our first point concerns the *observed* angular 2PCF. From the CDD of our total sample shown in Fig. 5 we have noted pairs in excess of random up to  $\theta_{max} = 30''$ . This is contradictory to Sesar et al. (2008) who find that there are essentially no physically bound pairs with an angular separation  $\theta > 15''$ . (We note, however, that this is in accord with some of our subsamples, see Fig. 13.) We considered several possible reasons for this apparent discrepancy, such as differences in the selection criteria of the Sesar et al. sample with respect to our sample, underestimation of the total area of holes in our sample, or an overcorrection of edge effects due to those holes, but we convinced ourselves that none of them can account for this disagreement. A real physical difference in the wide binary popula-



**Fig. 13.** 2PCF (solid circles) and CDD (open circles) for the subsamples A to H, together with the corresponding model curves. The subfigures are arranged as the corresponding subareas would be seen on the sky.



**Fig. 15.** 2PCF and CDD as inferred from our total sample as in Fig. 5. The model curve is for the broken power-law model from Lépine & Bongiorno (2007) (see text).

tion studied, e.g. caused by the fact that Sesar et al.'s sample is largely disc-dominated, while our final sample has a substantial halo contribution of  $\sim 30\%$ , can be excluded, as previous studies (Latham et al. 2002; Chanamé & Gould 2004) found that the disc and halo wide binary populations are reasonably consistent in their statistical characteristics.

Regarding the wide binary fraction, Sesar et al. also find a much smaller wide binary fraction of below 1% (decreasing with height above the Galactic plane), as compared to our  $f_{WB} \approx 10\%$ . Even if all pairs with a semi-major axis larger than 3 000 AU (beyond the "break") were ignored, we would still find a wide binary fraction of 4.6%. On the other hand, Lépine & Bongiorno

(2007) give a binary fraction of at least 9.5% for separations larger than 1 000 AU, which is in rough agreement with our results (setting  $a_{min} = 1 000$  AU we find  $f_{WB} \approx 8.3\%$ ).

Studying a much smaller region containing brighter stars as compared to our sample, WW87 and Garnavich (1988, 1991) found an unphysically large wide binary density in the direction of the NGP. Saarinen & Gilmore (1989) attribute this overdensity to a large statistical fluctuation. The region where this overdensity was found would be located in our subarea H. But as we probe fainter stars, no density enhancement is evident in that subarea. The slight overdensity we found in subarea E is different from that previously noted by WW87 and Garnavich.

As to the separation distribution, previous studies have found that the observational data are described by Öpik's law ( $\lambda = 1$ ) up to a certain maximum separation ("break"): Lépine & Bongiorno (2007) find this break to be around 3 500 AU, beyond which a steeper slope, with  $\lambda \approx 1.6$ , should apply. Sesar et al. (2008) basically agree with Lépine & Bongiorno and find in addition that the maximum separation increases with height above the Galactic plane. Poveda & Allen (2004) divided their wide binary catalogue (Poveda et al. 1994) into two subsamples consisting of the oldest and youngest systems, respectively, and find that the "maximum major semiaxis for which Öpik's distribution holds is much larger for the youngest binaries ( $a_m = 7862$  AU) than for the oldest ( $a_m = 2409$  AU)". Chanamé & Gould (2004) find the data to be described by a single powerlaw with an index of approximately 1.6 for their disc and halo samples. However, they note a "puzzling" flattening of the distribution of the disc binaries between  $10''$  and  $25''$ . As already suggested by Sesar et al., this flat range might be the domain where Öpik's law is valid. In view of the substantial uncertainties inherent in studies by Lépine & Bongiorno and by Chanamé & Gould, the two studies appear to be broadly consistent. Similar to our study, Garnavich (1991) assumed that the semi-major axis distribution is described by a single powerlaw. He finds the power-law index to be  $0.7 \pm 0.2$  for his NGP sample covering nearly 240 square degree. At intermediate Galactic latitudes he finds the slope to be steeper:  $\lambda = 1.3 \pm 0.2$ . His data

seem to favour a (somewhat unrealistic<sup>14</sup>) “cutoff” around 0.1 pc, especially for the NGP sample.

The question of a break in the separation distribution has much been discussed in the context of DM constraints. Wasserman & Weinberg (1991), for example, show that the observational data suggest a break but they do not *require* one statistically. Looking at the CDD for our total sample in Fig. 5, we note that our best-fit model slightly overestimates the number of pairs in excess of random at larger angular separations, and a more curved model line would be preferred by the data. This could indeed be interpreted as a hint that the semi-major axis distribution is broken, because having a steeper power-law index from a certain semi-major axis on would result in a flatter CDD model curve. In principle, we could easily fit a broken power-law to our data as well. However, too many free parameters will only destabilise our modelling. Instead we check whether our data are also consistent with the specific broken power-law distribution found by Lépine & Bongiorno (2007). To this end we use a broken reduced semi-major axis distribution of the form

$$q(a) = q_1(a) + q_2(a) = c_{\lambda_1} a^{-\lambda_1} \Theta(a_c - a) + c_{\lambda_2} a^{-\lambda_2} \Theta(a - a_c) \quad (48)$$

with the Heaviside function  $\Theta$  that introduces a break at semi-major axis  $a_c$ . It is normalised, as in §4.1.1, according to

$$\int_{a_{\min}}^{a_T} q(a) da = \int_{a_{\min}}^{a_c} q_1(a) da + \int_{a_c}^{a_T} q_2(a) da = 1. \quad (49)$$

Using the values given by Lépine & Bongiorno (2007),  $\lambda_1 = 1$ ,  $\lambda_2 = 1.6$ , and  $a_c = 0.02$  pc, we find for the normalisation constants

$$c_{\lambda_1} = \left[ \ln a_c - \ln a_{\min} + \frac{(a_T/a_c)^{1-\lambda_2} - 1}{1 - \lambda_2} \right]^{-1} \quad (50)$$

and

$$c_{\lambda_2} = c_{\lambda_1} a_c^{\lambda_2 - 1}. \quad (51)$$

The distribution of the projected separations  $Q(s)$  then splits into a sum, too:

$$Q(s) = \left( \frac{s}{pc} \right)^{-\lambda_1} C_{\lambda_1}(s) \text{ pc}^{-1} + \left( \frac{s}{pc} \right)^{-\lambda_2} C_{\lambda_2}(s) \text{ pc}^{-1}, \quad (52)$$

where  $C_{\lambda_i}(s)$  is defined as in Eq. 24 with the limits chosen appropriately when integrating over  $\eta$ .

We now keep the parameter given by Lépine & Bongiorno fixed so that our broken power-law model has only one free parameter: the wide binary density  $n_{WB}$ . In Fig. 15 we show the best-fit model curves corresponding to  $n_{WB} = 0.0075 \text{ pc}^{-3}$ . The broken power-law model also gives a decent fit to the 2PCF inferred from our total sample. (One should not be misled by large discrepancy between the model and the CDD showing the *cumulative* difference between the 2PCF and the corresponding model curve.) With a goodness-of-fit of  $Q \simeq 0.0047$  and a reduced chi-square of  $\chi^2/\nu \simeq 1.85$ , the broken power-law model provides a significantly worse bestfit to the data than our single power-law model. Nevertheless, as the errors in the 2PCF are slightly underestimated, the model is not put into question until  $Q < 0.001$  (Press et al. 1992, their §15), and we conclude that the broken power-law distribution found by Lépine & Bongiorno (2007) can not be rejected with confidence either.

<sup>14</sup> “Physically, one doesn’t expect to see a sharp cutoff.” (Weinberg 1990)

## 6.2. Internal (in)consistencies

Figure 10 shows that the best-fitting values for  $\lambda$  and  $n_{WB}$  are inconsistent, within the adopted *random* errors, when the limiting magnitude is varied. Both parameters are systematically shifted by roughly 20% when the limiting magnitude is changed by 1 mag. Further inconsistencies were met when we tried to differentiate our total sample with respect to colour (not shown here). This very likely means that there are unaccounted for *systematic* errors in the modelling, resting on oversimplified assumptions.

A basic assumption on the properties of wide binaries was that both components are drawn at random from the same stellar LF. There is, however, some evidence that this is not quite correct: Gould et al. (1995) noticed that “binaries are bluer and more distant than one would expect if they were formed of random combinations of field stars”. Also Lépine & Bongiorno (2007) find “that the luminosity function of the secondaries is significantly different from that of the single stars field population, showing a relative deficiency in low-luminosity ( $8 < M_V < 14$ ) objects”. Similarly, Sesar et al. (2008) report that blue stars ( $g - i \lesssim 2.0$  mag, corresponding roughly to  $g - r \lesssim 1.4$  mag) that are a member of a wide binary, have more blue companions than expected from the LF. (For red stars, however, they find that the components are drawn randomly from the LF.) Moreover, it has long been known (e.g. Bahcall & Soneira 1980, their Fig. 2) that stars of early spectral type have smaller scale heights than late type stars; therefore, the assumption that the stellar density distribution depends only on distance, and not on the luminosity of the stars, might be an oversimplification as well.

On the other hand, the internal inconsistency discussed is overemphasised in Fig. 10 because we only include the statistical errors stemming from the pair counts when determining the MCCR. If we also included the errors in the Galactic structure parameters and those from the LF, the MCCR would be considerably larger, and the inconsistency evident from Fig. 10 would not be as severe as it seems. That a variation in the Galactic structure parameter has a significant impact at least on  $n_{WB}$ , is apparent from Fig. 7. The systematic drift towards larger  $\lambda$  and  $n_{WB}$ , when adopting a brighter upper apparent magnitude limit (cf. Fig. 10), still points to some inconsistencies in our model. We think, however, that these inconsistencies do not entirely undermine our major results, although the quoted uncertainties might be considerably larger.

## 7. Summary and conclusions

We have derived the angular 2PCF for nearly 670 000 SDSS stars brighter than  $r = 20.5$  mag and redder than  $g - r = 0.5$  mag in a region of approximately 670 square degrees around the NGP. Various corrections had to be made for quasar contamination, survey holes, and bright stars. There is an unambiguous correlation signal on small scales up to  $30''$ . We modelled this signal by a modified WW-technique, closely following a previous study of Garnavich (1988, 1991) that was based on a much smaller sample of stars. The modelling involved a number of parametrised distribution functions: the spatial density and LF of stars, as well as the separation distribution of binaries. The Galactic model used is based on the recent study by Jurić et al. (2008) and the stellar LF derived by Jahreiß & Wielen (1997). For the wide binary semi-major axis distribution we assumed a single powerlaw. Furthermore, essential assumptions are: (1) binary stars follow the same density distribution as single stars, (2) both components of a binary are randomly drawn from the single star luminosity function. These assumptions allowed a significant

simplification of the model. The Galactic structure parameters were fixed (we explore three different sets), while the local wide binary number density  $n_{\text{WB}}$  and the power-law index  $\lambda$  of the semi-major axis distribution were left free; i.e., they have been determined by a least-squares fitting algorithm.

The best fit to the observed angular 2PCF of the total sample was obtained with  $\lambda \approx 1.0$ , which corresponds to the canonical Öpik law, and  $n_{\text{WB}} \approx 0.005 \text{ pc}^{-3}$ , meaning an overall local wide binary fraction of about 10.0% in the projected separation range of 0.001 pc (200 AU) to 1 pc (Galactic tidal limit). Previous studies (Poveda & Allen 2004; Lépine & Bongiorno 2007; Sesar et al. 2008) have also found the data to be consistent with Öpik's law, but only to a maximal separation that is considerably smaller than the Galactic tidal limit. Beyond that maximum separation, the distribution continues with a steeper decline. We have shown that our data are also consistent with the broken powerlaw found by Lépine & Bongiorno, although with a fit of lower quality (though involving only one free parameter, namely,  $n_{\text{WB}}$ ). Given this ambiguity, we are not able to put any strong constraints on the presence of a break in the wide binary separation distribution, which is regarded as one of the most interesting aspects of wide binaries. As to the wide binary fraction, we are in good accord with Lépine & Bongiorno (2007), whereas an apparent discrepancy with Sesar et al. (2008) remains.

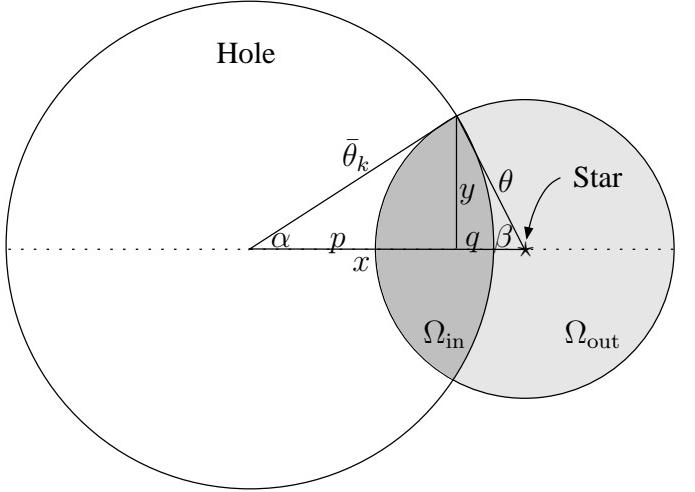
A differentiation of the sample with respect to limiting apparent magnitude turned up a systematic dependence of the binary parameters on the sample depth, which is most probably an artefact caused by oversimplified model assumptions. This conjecture is strengthened by the impossibility obtaining self-consistent results when differentiating with respect to colour. We conclude that one or more of the simplifying assumptions put into the model (e.g. that the luminosities of the binary components are independent of each other and draw randomly from the single-star LF or that binaries' density distribution follows exactly the single-star density distribution and that the density distribution is independent of the stars' luminosities) are not quite correct.

Differentiating the sample in terms of direction in the sky did yield some modest but non-significant variations. Only in one direction (subarea E) was an unexplained overdensity found.

While we have shown here that the stellar angular 2PCF, as a complement to common proper motion studies, basically works and remains a viable tool for the study of wide binaries, it has become clear that this method is severely limited by the need for – even more – complex modelling. To relax those simple model assumptions would unduly complicate the analysis much further and probably no longer yield unique solutions. Any more efficient progress will indeed have to involve distance information to discriminate against unwanted chance projections. We plan to include distance information in our future work. In spite of the limitations of the present, simple modelling of the angular 2PCF, we think that the general result for the total sample derived here, i.e.  $\lambda \approx 1$  (Öpik law) and  $f_{\text{WB}} \approx 10\%$  among stars having a spectral type later than G5, to within an uncertainty of 10%-20%, is a robust result.

## Appendix A: Edge correction for holes

To quantify the edge effect from holes in our sample, we assume that the holes are circular and flat (Euclidian) and that effects due



**Fig. A.1.** Illustration of the geometry used to correct for edge effects due to holes. The large circle represents the bounding circle of a hole or a bright star mask. A star lying close to the hole is indicated. The region within a distance  $\theta$  of that star (highlighted in grey) is partially inside the hole (dark grey intersection).

to intersecting holes are negligible. Let  $\bar{A}_k(\theta)$  be the solid angle of the annulus of width  $2\theta$  around a hole of radius  $\bar{\theta}_k$

$$\bar{A}_k(\theta) = 4\pi\theta(\bar{\theta}_k + \theta), \quad \text{with } k = \{\text{SH, BS}\}, \quad (\text{A.1})$$

where  $k$  stands for hole masks (survey holes: SH) or bright star masks (BS). Due to the holes in our sample we observe only a fraction  $F_H^{\text{tot}}(\theta)$  of all stellar pairs in the residual solid angle  $\Omega$  sparated by an angular distance  $\theta$

$$F_H^{\text{tot}}(\theta) = 1 - \frac{1}{\Omega} \sum_{k=\{\text{SH, BS}\}} N_k \bar{A}_k(\theta) (1 - F_H(\theta; \bar{\theta}_k)), \quad (\text{A.2})$$

where  $F_H(\theta; \bar{\theta}_k)$  is the fraction of all pairs separated by  $\theta$  we observe in  $A_k(\theta)$

$$F_H(\theta; \bar{\theta}_k) = \frac{2\pi}{\bar{A}_k(\theta)} \int_{\bar{\theta}_k}^{\bar{\theta}_k+\theta} xf(x; \theta, \bar{\theta}_k) dx. \quad (\text{A.3})$$

Here,  $f$  is the fraction of the area  $\pi\theta^2$  of the disc with radius  $\theta$  around a star that lies outside the hole (Fig. A.1)

$$f(x; \theta, \bar{\theta}_k) = \frac{\Omega_{\text{out}}(x; \theta, \bar{\theta}_k)}{\pi\theta^2} = 1 - \frac{\Omega_{\text{in}}(x; \theta, \bar{\theta}_k)}{\pi\theta^2}, \quad (\text{A.4})$$

whereas  $\Omega_{\text{in}} = \pi\theta^2 - \Omega_{\text{out}}$  is the area of that disc within the hole, i.e. the intersection. With the nomenclature given in Fig. A.1, we find

$$\Omega_{\text{out}} = \bar{\theta}_k^2 \alpha - yp + \theta^2 \beta - yq, \quad (\text{A.5})$$

where

$$\alpha = \arcsin\left(\frac{y}{\bar{\theta}_k}\right); \quad \beta = \arcsin\left(\frac{y}{\theta}\right) \quad (\text{A.6})$$

and

$$p = \sqrt{\bar{\theta}_k^2 - y^2}; \quad q = \sqrt{\theta^2 - y^2}. \quad (\text{A.7})$$

Solving  $x = p + q$  for  $y$  using *Maple* gives

$$y = \frac{1}{2x} \sqrt{2\bar{\theta}_k^2 x^2 - x^4 - \theta^4 + 2\theta^2 x^2 + 2\theta^2 \bar{\theta}_k^2 - \bar{\theta}_k^4}, \quad (\text{A.8})$$

where the integration in (A.3) was performed with *Maple*, too.

## Appendix B: Galactic tidal limit

A crude estimate of the Galactic tidal limit  $a_T$ , i.e. of the maximum semi-major axis  $a_{\max}$  of a binary star with total mass  $M$  orbiting in the Galactic tidal field is provided by the *Jacobi limit*  $r_J$  (sometimes also referred to as *zero-velocity surface* or *Roche sphere*) (e.g. Binney & Tremaine 2008, §8.3)

$$r_J \simeq \left( \frac{M}{3M(<\tilde{D})} \right)^{\frac{1}{3}} \tilde{D}, \quad (\text{B.1})$$

where  $\tilde{D}$  is the galactocentric distance of the binary system and  $M(<\tilde{D})$  is the Galactic mass enclosed within radius  $\tilde{D}$ . Then,  $M(<\tilde{D})$  is given by

$$M(<\tilde{D}) = \frac{v_c^2}{G} \tilde{D} \simeq 1.1 \cdot 10^7 \left( \frac{\tilde{D}}{\text{pc}} \right) M_\odot, \quad (\text{B.2})$$

where we have adopted the canonical value of the circular speed  $v_c = 220 \text{ km s}^{-1}$ .

For a rough estimate of this distance range, it is useful to define the *effective depth* of the sample as the most likely distance of an arbitrarily chosen star (Weinberg & Wasserman 1988). In a magnitude-limited sample the total number of stars expected to be observed in a (heliocentric) distance range between  $D$  and  $D + dD$  is

$$dN(D) = \Omega dDD^2 \rho(D) \int_{M_{\min}(D)}^{M_{\max}(D)} dM \Phi(M), \quad (\text{B.3})$$

where  $\Omega$  denotes the solid angle covered by our sample,  $\rho(D)$  the density distribution, and  $\Phi(M)$  the stellar LF. The absolute magnitudes  $M_{\min}(D)$  and  $M_{\max}(D)$  are given by Eq. 27 and 28, respectively. The effective depth  $D_{\text{eff}}$  is then given by the median value of the distribution  $dN(D)$  (Weinberg & Wasserman 1988)

$$\frac{1}{2} = \frac{\int_0^{D_{\text{eff}}} dN(D)}{\int_0^\infty dN(D)} = \frac{\int_{D_{\min}}^{D_{\text{eff}}} dN(D)}{\int_{D_{\min}}^{D_{\max}} dN(D)}, \quad (\text{B.4})$$

where  $D_{\min}$  and  $D_{\max}$  are used for the numerical integration and are chosen to bracket the theoretical distance range probed by our sample, which – neglecting interstellar dust extinction – is readily found from the apparent magnitude limits,  $m_{\min}$  and  $m_{\max}$ , and the absolute magnitude limits,  $M_{\min}$  and  $M_{\max}$ , given by  $\Phi(M)$ . We take them to be

$$D_{\min} = 10^{(m_{\min} - M_{\max} + 5)/5} = 10^{(15 - 21.5 + 5)/5} \simeq 0.5 \text{ pc} \quad (\text{B.5})$$

and

$$D_{\max} = 10^{(m_{\max} - M_{\min} + 5)/5} = 10^{(20.5 - (-0.5) + 5)/5} \simeq 160 \text{ kpc}. \quad (\text{B.6})$$

In Table 2 we list the values of  $D_{\text{eff}}$  for our various (sub)samples calculated using the Galactic structure parameter set 2. With the notation of §4.2.1, the effective galactocentric distance  $\tilde{D}_{\text{eff}}$  can be expressed as

$$\begin{aligned} \tilde{D}_{\text{eff}}^2 = & r_0^2 + D_{\text{eff}}^2 \cos^2 b - 2r_0 D_{\text{eff}} \cos b \cos \ell \\ & + z_0^2 + 2z_0 D_{\text{eff}} \sin b + D_{\text{eff}}^2 \sin^2 b. \end{aligned} \quad (\text{B.7})$$

Its values are typically somewhat above 8 kpc.

The last parameter we need to specify in order to estimate the Galactic tidal limit  $a_T$ , is the total mass of the binary star

$M$ . Given the magnitude limits of our sample we may write the average mass of a star in the sample as

$$\langle m \rangle = \frac{\int_{m_{\min}}^{m_{\max}} m \xi(m) dm}{\int_{m_{\min}}^{m_{\max}} \xi(m) dm} = \frac{\int_{M_{\min}}^{M_{\max}} m(M) \Phi(M) dM}{\int_{M_{\min}}^{M_{\max}} \Phi(M) dM}, \quad (\text{B.8})$$

where  $\xi(m)$  denotes the mass function and  $m(M)$  the mass-luminosity relation which we take from Kroupa et al. (1993). The transformation into  $r$  magnitudes was performed as in §4.2.2.

Statistically, we expect the total mass of a binary star to be  $M = 2\langle m \rangle$ , and we finally estimate the Galactic tidal limit according to

$$\begin{aligned} a_T(\tilde{D}_{\text{eff}}, M) \simeq & \left( \frac{M}{3M(<\tilde{D}_{\text{eff}})} \right)^{\frac{1}{3}} \tilde{D}_{\text{eff}} \\ & \simeq 3.1 \cdot 10^{-3} \left( \frac{M}{M_\odot} \right)^{\frac{1}{3}} \left( \frac{\tilde{D}_{\text{eff}}}{\text{pc}} \right)^{\frac{2}{3}} \text{ pc}. \end{aligned} \quad (\text{B.9})$$

The values of  $a_T$  are around 1 pc and we list them for various (sub)samples in Table 2.

*Acknowledgements.* We are grateful to P. M. Garnavich for providing us with his PhD thesis, which served as a starting point for our study. It is a pleasure to thank R. Buser, B. Krusche, T. Lisker, N. Sambhus, I. Sick, G. A. Tammann, C. von Arx, A. Willand, and T. Zingg for helpful comments and valuable discussions of the present work. We are much obliged to the referee for a prompt and insightful report. We are indebted to N. Peduzzi and the A&A language editor J. Adams for proofreading our manuscript.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

This work was supported by the Swiss National Science Foundation.

## References

- Abt, H. A. 1983, ARA&A, 21, 343
- Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2008, ApJS, 175, 297
- Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2007, ApJS, 172, 634
- Allen, C., Poveda, A., & Hernández-Alcántara, A. 2007, in IAU Symposium, Vol. 240, IAU Symposium, ed. W. I. Hartkopf, E. F. Guinan, & P. Harmancic, 405–413
- Bahcall, J. N., Hut, P., & Tremaine, S. 1985a, ApJ, 290, 15
- Bahcall, J. N., Ratnatunga, K. U., Buser, R., Fenkart, R. P., & Spaenhauer, A. 1985b, ApJ, 299, 616
- Bahcall, J. N. & Soneira, R. M. 1980, ApJS, 44, 73
- Bahcall, J. N. & Soneira, R. M. 1981, ApJ, 246, 122
- Bernstein, G. M. 1994, ApJ, 424, 569
- Bilir, S., Karaali, S., & Tunçel, S. 2005, Astronomische Nachrichten, 326, 321

- Binney, J. & Tremaine, S. 2008, Galactic Dynamics: Second Edition, ed. J. Binney & S. Tremaine (Princeton University Press)
- Chamamé, J. 2007, in IAU Symposium, Vol. 240, IAU Symposium, ed. W. I. Hartkopf, E. F. Guinan, & P. Harmanec, 316–325
- Chamamé, J. & Gould, A. 2004, ApJ, 601, 289
- Chen, B., Stoughton, C., Smith, J. A., et al. 2001, ApJ, 553, 184
- Close, L. M., Richer, H. B., & Crabtree, D. R. 1990, AJ, 100, 1968
- Doinidis, S. P. & Beers, T. C. 1989, ApJ, 340, L57
- Finlator, K., Ivezić, Ž., Fan, X., et al. 2000, AJ, 120, 2615
- Garnavich, P. M. 1988, ApJ, 335, L47
- Garnavich, P. M. 1991, PhD thesis, AA(Washington Univ., Saint Louis, MO.)
- Gomez, M. & Lada, C. J. 1998, AJ, 115, 1524
- Goodwin, S. P., Kroupa, P., Goodman, A., & Burkert, A. 2007, in Protostars and Planets V, ed. B. Reipurth, D. Jewitt, & K. Keil, 133–147
- Gould, A., Bahcall, J. N., Maoz, D., & Yanny, B. 1995, ApJ, 441, 200
- Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040
- Hamilton, A. J. S. 1993, ApJ, 417, 19
- Harris, H. C., Munn, J. A., Kilic, M., et al. 2006, AJ, 131, 571
- Hernandez, X. & Lee, W. H. 2008, MNRAS, 387, 1727
- Jahreiß, H. & Wielen, R. 1997, in ESA Special Publication, Vol. 402, Hipparcos – Venice '97, 675–680
- Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864
- Karttunen, H., Kröger, P., Oja, H., Poutanen, M., & Donner, K. J. 1996, Fundamental Astronomy, ed. H. Karttunen, P. Kröger, H. Oja, M. Poutanen, & K. J. Donner
- Kobulnicky, H. A. & Fryer, C. L. 2007, ApJ, 670, 747
- Kouwenhoven, M. B. N., Goodwin, S. P., Parker, R. J., et al. 2010, ArXiv e-prints
- Kroupa, P., Tout, C. A., & Gilmore, G. 1993, MNRAS, 262, 545
- Laird, J. B., Carney, B. W., & Latham, D. W. 1988, AJ, 95, 1843
- Larson, R. B. 2001, in IAU Symposium, Vol. 200, The Formation of Binary Stars, ed. H. Zinnecker & R. Mathieu, 93–104
- Latham, D. W., Schechter, P., Tonry, J., Bahcall, J. N., & Soneira, R. M. 1984, ApJ, 281, L41
- Latham, D. W., Stefanik, R. P., Torres, G., et al. 2002, AJ, 124, 1144
- Lépine, S. & Bongiorno, B. 2007, AJ, 133, 889
- López-Corredoira, M., Garzón, F., Hammersley, P. L., & Mahoney, T. J. 1998, MNRAS, 301, 289
- Lourakis, M. 2004, levmar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++, <http://www.ics.forth.gr/~lourakis/levmar/>
- Lupton, R., Gunn, J. E., Ivezić, Ž., Knapp, G. R., & Kent, S. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 238, Astronomical Data Analysis Software and Systems X, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne, 269–278
- Milgrom, M. & Bekenstein, J. 1987, in IAU Symposium, Vol. 117, Dark matter in the universe, ed. J. Kormendy & G. R. Knapp, 319–330
- Öpik, E. 1924, Publications of the Tartu Astrofizika Observatory, 25, 1
- Parker, R. J., Goodwin, S. P., Kroupa, P., & Kouwenhoven, M. B. N. 2009, MNRAS, 873
- Peebles, P. J. E. 1980, The large-scale structure of the universe (Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p.)
- Poveda, A. & Allen, C. 2004, in Revista Mexicana de Astronomía y Astrofísica Conference Series, Vol. 21, Revista Mexicana de Astronomía y Astrofísica Conference Series, ed. C. Allen & C. Scarfe, 49–57
- Poveda, A., Allen, C., & Hernández-Alcántara, A. 2007, in IAU Symposium, Vol. 240, IAU Symposium, ed. W. I. Hartkopf, E. F. Guinan, & P. Harmanec, 417–425
- Poveda, A., Herrera, M. A., Allen, C., Cordero, G., & Lavalley, C. 1994, Revista Mexicana de Astronomía y Astrofísica, 28, 43
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical recipes in C. The art of scientific computing (Cambridge: University Press, —c1992, 2nd ed.)
- Quinn, D. P., Wilkinson, M. I., Irwin, M. J., et al. 2009, MNRAS, 396, L11
- Reid, I. N. & Gizis, J. E. 1997, AJ, 113, 2246
- Retterer, J. M. & King, I. R. 1982, ApJ, 254, 214
- Saarinen, S. & Gilmore, G. 1989, MNRAS, 237, 311
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
- Scratton, R., Johnston, D., Dodelson, S., et al. 2002, ApJ, 579, 48
- Sesar, B., Ivezić, Ž., & Jurić, M. 2008, ApJ, 689, 1244
- Valtonen, M. 1997, Celestial Mechanics and Dynamical Astronomy, 68, 27
- Wasserman, I. & Weinberg, M. D. 1987, ApJ, 312, 390 (WW87)
- Wasserman, I. & Weinberg, M. D. 1991, ApJ, 382, 149
- Weinberg, M. D. 1988, Ap&SS, 142, 277
- Weinberg, M. D. 1990, in NATO ASIC Proc. 305: Baryonic Dark Matter, 117–136
- Weinberg, M. D., Shapiro, S. L., & Wasserman, I. 1987, ApJ, 312, 367
- Weinberg, M. D. & Wasserman, I. 1988, ApJ, 329, 253
- Yoo, J., Chamamé, J., & Gould, A. 2004, ApJ, 601, 311
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579